AL.2.1986-490

PLANNING FOR AN EVALUATION OF TEACHING PERFORMANCE

CANADIANA
C✓
FEB 24 1986

VOLUME IV

Summaries Of Instruments For Use in Evaluating Teaching Performance

Planning Services

Alberta
EDUCATION

DDN 6273354

[redacted]

PLANNING FOR AN EVALUATION OF TEACHING PERFORMANCE

CANADIANA
C2
FEB 24 1986

VOLUME IV

Summaries Of Instruments For Use in Evaluating Teaching Performance

Compiled by
I. Yuzdepski

With
L. Elliott

Planning Services

Alberta Education

August, 1985

Digitized by the Internet Archive
in 2015

https://archive.org/details/planningforevalu04yuzd

FOREWORD

Planning for an Evaluation of Teaching Performance comprises four separately bound volumes: A Manual of Guidelines, (Vol. I); Teacher Effectiveness: Annotated Bibliography and References, (Vol. II); Evaluation Procedures: Annotated Bibliography and References, (Vol. III); and Summaries of Instruments for Use in Evaluating Teaching Performance, (Vol. IV).

The content is delimited to gathering and interpreting information about teaching and learning in classroom settings. The main focus is on behaviors of teachers that research has shown to be consistently associated with academic learning by pupils. Some attention is also directed to several classroom behaviors of pupils often found to correlate with effective learning.

The contents of the publication are intended to be used as guidelines. Readers should feel free to reject, adapt or adopt the ideas communicated as local needs dictate. All school jurisdictions have policy and structures in place for evaluating teaching performance. Hence, selections from these volumes might be identified and used in reviewing local policies, guidelines and procedures.

The companion publication, Evaluating and Improving Teaching Performance: An Inservice Kit, by Dr. Laurie Mireau is designed primarily for inservice settings. The Inservice Kit is designed for "hands-on" experience by individuals and by groups at the school and district levels. The Kit illustrates and emphasizes practical considerations and procedures useful in evaluating classroom activities and for integrating these with supervision of instruction.

An integral part of the Kit is a set of 30 videotapes (14 training tapes and 16 demonstration tapes) involving eleven classroom teachers in Edmonton Public Schools. The reader is directed to A Manual of Guidelines, (Vol. I), Appendix 2, for detailed descriptions of the 30 videotapes.

VOLUME IV

SUMMARIES OF INSTRUMENTS FOR USE IN EVALUATING TEACHING PERFORMANCE

The purpose of Volume IV is to assist personnel involved in the evaluation process, identify observation instruments which may be useful in preparing appropriate evaluation instruments in their areas. Few, if any, of the instruments for use by observers in this volume will be appropriate for direct use in their present form for local evaluations. However, descriptors and terms from the descriptions about the instruments may offer guidance in developing recording form(s) for use locally.

This volume includes:

1. **A chart** listing the nineteen criteria and their descriptions. These were introduced in A Manual of Guidelines, (Vol. 1), Chapter 4.

2. **Summary sheets** of 54 evaluation instruments. The first set of instruments (43) is concerned with the observation and recording of teaching behaviors on rating scales or checklists by a supervisor (Section A). Eleven instruments for recording student perceptions of teaching performance comprise the second set of instruments (Section B).

   Each summary includes pertinent information about the instrument regarding publishing company, author, criteria evaluated, subject of observation, category dimension, and coding units.

   A more detailed description of the instrument will follow those published in Evaluating Classroom Instruction, A Sourcebook of Instruments (Borich and Madden, 1977).

3. **Reference tables** which indicate what particular instruments might be useful in observing specific criteria. Table 1 depicts 43 instruments which may be used by a supervisor; Table 2 refers to 11 instruments for pupil assessment of teaching performance.

   Abbreviated instrument names are listed alphabetically down the left side of the table, and the numbers 1 to 19 along the top refer to the 19 criteria. The "X" indicates which of the 19 criteria can be recorded on the instrument. Again the reader is cautioned that the nineteen criteria are not "standardized". The descriptions of the criteria are a more valid representation of the behavior than the "label" which others may use in slightly different ways.

NOTE: For those readers unfamiliar with observation methods and types of observation instruments, it is recommended that Chapters 5 and 6 in A Manual of Guidelines be read before this Volume.

NOTE: The percentages of agreement between two observers are frequently used as indices of reliability. The computation formula is as follows:

> % of agreement = $100 \times (1 - \frac{x - y}{x + y})$ where x = the number of the particular target behavior recorded by the first observer or during the first observation by a single observer, and y = the number recorded by the second observer or during the second observation by a single observer (two separated observations of the same type by the same single observer are used to establish intra-observer reliability).

A lower limit within the range of 70-85 percent agreement is often set. If this standard is not met the decision will be to reject or revise the instrument and/or undertake more training of observer-recorders.

Chart A

Criteria Numbers	Criteria Names	Criteria Descriptions
1.	Plans, Organization, and Implementation	Arrangements and procedures worked out in advance of instruction by the teacher and the extent to which these are appropriate and used.
2.	Academic Learning Time	The time a student is engaged with instructional materials or activities that allow a high level of success for that student.
3.	Teacher's Expectations	The appropriateness of the teacher's predictions of the individual's or group's achievements in a particular content area.
4.	Teacher's Awareness	Alertness to pupil behavior at all times; "withitness," having "eyes in the back of one's head" (Kounin, 1970).
5.	Concurrent Tasks Management	The efficiency and ease with which a teacher handles various classroom activities at the same time.
6.	Pace of Instruction	The rate at which learning experiences are provided for pupils by the teacher.
7.	Classroom Management	The teacher's facility to anticipate, prevent and deal satisfactorily with pupils' interpersonal and academic difficulties. Success in appropriately disciplining pupils.

Criteria Numbers	Criteria Names	Criteria Descriptions
8.	Feedback to Pupils	Verbal and non-verbal responses to students.
9.	Time on Task	Engaged time; the time a student is attending to instruction in a particular content area.
10.	Enthusiasm	Teacher's display of eagerness and zeal.
11.	Clarity of Instruction	The communication of the teacher is such that the students are totally informed about what they should be doing, where to do it, and for how long, in order to ensure that students do not lose time because they are confused or because they are waiting for decisions to be made.
12.	Task Oriented/ Businesslike Orientation	The degree of academic focus and the teacher's directness in providing academically focussed instruction.
13.	Variability of Instruction	The appropriate use of variety in providing learning experiences for pupils.
14.	Flexibility	The ease with which the teacher adapts and adjusts satisfactorily to varying factors in the classroom setting.
15.	Understanding of Instructional Goals	The pupils' comprehension about what they should be doing, where to do it, and for how long.

Criteria Numbers	Criteria Names	Criteria Descriptions
16.	Teacher Centred/Direct Instruction	"An instructional strategy where time spent on instruction is sufficient and continuous, content covered is extensive, goals are clear to students, materials are structured and sequential, questions are low level to produce many correct student responses, student progress is monitored, feedback is immediate and correct responses are reinforced. With direct instruction, the teacher sets the goals, plans the activities, chooses the material; questions are convergent, not divergent; interaction is structured, not authoritarian; tasks are cognitive oriented, not humanistic" (Ornstein, 1984: 343).
17.	Praise and Encouragement/Criticism	The appropriate use of positive (e.g., praise and encouragement) and negative (e.g., criticism) reinforcement methods.
18.	Individualization of Instruction	Teaching adapted to the background, ability, and needs of individual learners.
19.	Questioning Techniques	Matching the types of questions asked of the student to the characteristics of the student and to the instructional objectives.

Section A:

Instrument Summaries for Supervisor Use

Table 1

Instruments and Criteria for Supervisor Use

		Categories																		
	Instrument	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
1.	Cl. Beh. Tape Anal.	X							X					X				X		X
2.	Cl. Communication							X	X								X	X		X
3.	Cl. Obs. Code Digest													X			X		X	
4.	Cl. Obs. Record			X				X						X			X			
5.	Classroom Creativity										X			X				X		
6.	Classroom Obs. Scales									X	X	X			X	X	X	X		X
7.	Cognitive Components			X										X						X
8.	Connors-Eisenberg	X										X		X				X		
9.	CASES									X						X				
10.	CERLI Verbal-Beh.													X						X
11.	Disruptive Behaviour									X										
12.	Dodl													X			X	X		X
13.	Effective Teacher							X			X							X		
14.	Florida Climate	X						X					X				X	X		
15.	FAIR 33							X							X			X		X
16.	Hough								X	X								X	X	
17.	Jackson, Lahaderne							X		X			X			X		X		
18.	Joyce			X					X					X	X			X	X	X
19.	Kounin				X	X		X		X										
20.	Kounin, Friesen				X	X		X		X		X			X			X		
21.	MacDonald-Zaret			X					X						X		X	X		X
22.	Microenvironments							X		X							X	X		
23.	Miller-Hughes	X		X					X						X		X	X		
24.	Mini Total Interact.							X	X		X			X		X	X			
25.	Modified Category								X									X		X
26.	Nonverbal Interaction									X				X			X	X		X
27.	OSCAR 4V								X				X				X	X		X
28.	Provo Code	X		X					X									X		
29.	Pupil-Teacher Rapport									X			X				X	X		
30.	PART													X						
31.	Rev. Verb. Interact								X								X	X		X
32.	Social Weather												X		X		X	X		
33.	Student Activity									X										
34.	Student Behaviour									X										
35.	Student Obs. Form									X									X	
36.	SAVI							X							X		X	X		
37.	SICC														X		X	X		X
38.	STARS	X						X					X				X	X		
39.	Teacher Behaviour				X												X	X		X
40.	Teacher Verb-Nonverb				X							X	X				X	X		X
41.	Teacher-Child Dyadic			X					X									X		X
42.	Waimon								X			X					X	X		X
43.	Wallen's Flanders Mod.				X			X				X						X		X

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Classroom Behavior Tape Analysis Form (Cl. Beh. Tape Anal.#1)

AUTHOR: Daniel Solomon

AUTHOR/PUBLISHER ADDRESS: Center for the Study of Liberal Education for Adults
Chicago, Illinois

CRITERIA EVALUATED: 1, 8, 13, 17, 19

SUBJECT OF OBSERVATION

[✓ TEACHER AND PUPILS]
TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

1 ONLY
[✓ MORE THAN 2 PEOPLE IN CLASSROOM]

COLLECTION METHODS

LIVE
LIVE WITH CODING EQUIPMENT
[✓ VIDEO AND/OR AUDIO TAPE]

CATEGORY DIMENSIONS

[✓ AFFECTIVE]
[✓ COGNITIVE]
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

[✓ CATEGORY CHANGE]
TOPIC CHANGE

SPEAKER CHANGE
TIME UNIT
TIME SAMPLE
RATING SCALE
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

[✓ RESEARCH]
TRAINING
EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 150-152.

Name: CLASSROOM BEHAVIOR TAPE ANALYSIS FORM

Author: Daniel Solomon

Type of Measure: Observation coding system

Availability: Solomon, D. *Teaching styles and learning.* Chicago: Center for the Study of Liberal Education for Adults, 1963.

Description: The Classroom Behavior Tape Analysis Form (CBTAF) employs tape recordings of classroom interaction to yield indices of teacher and student behavior. Developed in a study of teacher style and student outcomes (Solomon, 1963), the CBTAF requires raters to tape and then classify verbal behaviors occurring in the classroom, using categories derived from: (1) a review of the literature; (2) theoretical and intuitive expectations; and (3) the results of an exploratory study by Solomon and Miller (1961). These categories, listed below, encompass 3 aspects of teacher behavior and 2 aspects of student participation.

I. Teacher Behavior
  A. Statements
    1. Providing Organization, Orientation
    2. Providing Hypothesis
    3. Providing Opinion
    4. Providing Factual
    5. Providing Interpretation
    6. Providing Personal Reference
  B. Questions
    7. Requesting Organization, Orientation
    8. Requesting Hypothesis
    9. Requesting Opinion
    10. Requesting Factual
    11. Requesting Interpretation
    12. Requesting, Nonspecific
    13. Requesting Clarification
  C. Feedback
    14. Positive Reinforcement
    15. Negative Reinforcement
    16. No Reinforcement
    17. High Information
    18. Low Information
    19. No Information
    20. Residual

II. Student Behavior
  A. Statements
    21. Providing Organization, Orientation
    22. Providing Hypothesis
    23. Providing Opinion
    24. Providing Factual
    25. Providing Interpretation
    26. Providing Personal Reference
  B. Questions
    27. Requesting Organization, Orientation
    28. Requesting Hypothesis
    29. Requesting Opinion
    30. Requesting Factual
    31. Providing Interpretation
    32. Providing, Nonspecific
    33. Providing Clarification
    34. Residual
    35. Tension Release

Sample Items: Solomon (1963) provides examples of the kinds of behaviors classified under each CBTAF category. "Requesting, Nonspecific," for example, includes any "unstructured invitation to participate (p. 138)," any comment which encourages group discussion. "No Reinforcement" refers to the teacher's "dispassionate reiteration of student's communication, silence, shift to another topic (p. 138)."

Reliability: Interrater reliability of the Tape Analysis Form was checked by Solomon (1963) on two occasions—at the end of the rater training period and midway through the actual tape analyses. On each of these occasions, 4 raters classified approximately 45 minutes of a tape to which they had not yet listened. The categories recorded by each rater were ranked according to their relative frequencies. The similarity among these rankings was computed using Kendall's coefficient of concordance. The first reliability check yielded a coefficient of .97 and the second, .84, each significant beyond the .001 level.

Intervisit correlations were also computed for each CBTAF category. The majority of these were above .50, and many exceeded .60, indicating a fair degree of stability in teacher behavior across class sessions (Solomon, 1963).

Validity: Solomon (1963) investigated concurrent validity by correlating CBTAF dimensions with similar variables measured by other instruments employed in his study. Amount of Teacher Speech on the CBTAF correlated significantly ($p < .05$) with the following items on the Student Questionnaire: Teacher Provides Main Directing Force (.80); Amount of Teacher Lecturing (.48); Instructor's Loss of Control (−.52); and Inter-Student Discussion (−.62). The CBTAF variable, Teacher's Personal References, correlated .54 with Expressed Opinions About Material and .65 with References to Self. Similarly, Teacher's Use of Negative Reinforcement was positively related (.71) to the Student Questionnaire item, Teacher's Use of Challenging, Embarrassing Techniques. Comparable correlations between CBTAF scores and observers' ratings of teacher behavior provide further evidence of validity.

Norms: The sample used in the Solomon (1963) study was composed of 24 college instructors teaching evening courses in Introductory American Government at 13 schools. All of the schools were located in the Midwest, except 1, which was in a large Eastern city. Class size ranged from 11 to 38. Students were predominantly male, "working full time," and from 16 to 60 years old (median age = 20).

Procedures for Use: In using the CBTAF, the entire class session is recorded. Statements and questions of the teacher and students are coded as they occur on the tape, all student speech being scored together rather than separately for individual students.

A number of indices can be derived from the tape, including: (1) the proportion of teacher and student talk falling into each of the speech categories; (2)

the ratio of each category of teacher talk to the same category of student talk; and (3) the percentage of total talk contributed by the teacher.

Comments: The Classroom Behavior Tape Analysis Form attempts to assign every statement made in the classroom to a meaningful category. The CBTAF focuses on relatively small units of behavior, which are scored immediately. The tape recording provides a permanent record of classroom activity; thus there is no reliance on the observer's memory. Because this instrument considers only the verbal behavior of students and teachers, it is perhaps best used in conjunction with other measures of classroom interaction. This is, in fact, how Solomon has employed the CBTAF in the past.

References: Solomon, D., *Teaching styles and learning.* Chicago: Center for the Study of Liberal Education for Adults, 1963.

Solomon, D., W. E. Bezdek, Rosenberg, L., "Dimensions of teacher behavior." *Journal of Experimental Education*, 1964, 33(1), 23–40.

Solomon, D., and H. L. Miller, *Exploration in teaching styles.* Chicago: Center for the Study of Liberal Education for Adults, 1961.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: System for the Analysis of Classroom Communication (Cl. Communication #2)

AUTHOR: Margaret Hubbard Jones
AUTHOR/PUBLISHER ADDRESS: University of California, Los Angeles, California

CRITERIA EVALUATED: 7, 8, 16, 17, 19

SUBJECT OF OBSERVATION

__x__ TEACHER AND PUPILS
_____ TEACHER ONLY
_____ PUPIL ONLY
_____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
_____ 1 ONLY
__x__ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

__x__ LIVE
_____ LIVE WITH CODING EQUIPMENT
_____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
__x__ AFFECTIVE
__x__ COGNITIVE
_____ PROCEDURE OR ROUTINE
_____ ACTIVITY
_____ SOCIOLOGICAL STRUCTURE
_____ OTHER

SETTINGS

__x__ CLASSROOM, ANY SUBJECT
_____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

__x__ CATEGORY CHANGE
_____ TOPIC CHANGE

__x__ SPEAKER CHANGE
__x__ TIME UNIT
_____ TIME SAMPLE
_____ RATING SCALE
_____ OTHER

OBSERVER(S) REQUIRED

__x__ 1 PERSON
_____ TEAM OF 2 PERSONS
_____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
__x__ RESEARCH
_____ TRAINING
__x__ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 333-335.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Classroom Observation Code Digest (Cl. Obs. Code Digest, #3)

AUTHOR: Francis G. Cornell, Carl L. Lindvall, Joe L. Saupe
AUTHOR/PUBLISHER ADDRESS: Bureau of Educational Research, College of Education

CRITERIA EVALUATED: 13, 16, 18

SUBJECT OF OBSERVATION

_x_ TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
___ 1 ONLY
_x_ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

_x_ LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
_x_ AFFECTIVE
_x_ COGNITIVE
_x_ PROCEDURE OR ROUTINE
_x_ ACTIVITY
_x_ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

_x_ CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

_x_ CATEGORY CHANGE
___ TOPIC CHANGE

___ SPEAKER CHANGE
___ TIME UNIT
_x_ TIME SAMPLE
_x_ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

_x_ 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
_x_ RESEARCH
___ TRAINING
___ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 438-440.

Name: CLASSROOM OBSERVATION CODE DIGEST

Authors: Francis G. Cornell, Carl M. Lindvall, and Joe L. Saupe

Type of Measure: Observation coding system

Availability: Cornell, F. G., C. M. Lindvall, and J. L. Saupe, *An exploratory measurement of individualities of schools and classrooms.* Urbana, Illinois: Bureau of Educational Research, College of Education, University of Illinois, 1952.

Description: The Classroom Observation Code Digest was developed to measure differences in classrooms as an index of differences in school systems. After an extensive review of the literature, an evaluation of related measurement attempts, and empirical testing, the following dimensions were selected for inclusion on the Digest:

1. Differentiation—the extent to which the teacher recognizes individual differences among students.
2. Social Organization—the group structure and the pattern of interaction characterizing the classroom.
3. Pupil Initiative—the extent to which pupils are allowed to determine the learning situation.
4. Content—what is taught in the classroom.
5. Variety—the degree to which a variety of activities or techniques are used.
6. Competency—the technical performance of the teacher.
7. Classroom Climate (Teacher)—the social/emotional atmosphere of the classroom as reflected in the teacher's behavior.
8. Classroom Climate (Pupils)—the social/emotional atmosphere of the classroom as reflected in the pupils' behavior.

Sample Items: Each of the 8 dimensions represented on the Digest is designed to measure 1 aspect of a classroom learning situation. The Digest contains brief definitions of behavioral codes which are recorded by the observer on an accompanying tally sheet, labeled the Classroom Observation Schedule. The specific behaviors listed below are similar to those included in several of the Digest categories:

Dimension	*Behavior*
Variety	1. Pupils involved in class discussion. 2. Pupils performed a demonstration. 3. Teacher read aloud from text. 4. Teacher showed a film.

Competency
1. Teacher conducted a review lesson.
2. Teacher provided relevant examples.
3. Teacher allowed discussion to continue aimlessly.
4. Teacher's answers to pupil questions were vague and ambiguous.

Reliability: In order to determine the instrument's reliability, the Digest was tested in 32 classrooms, each visited 3 times by 6 observers in teams of 2. During each visit, observers independently recorded behaviors for 30 minutes. Coefficients reported by Cornell, Lindvall, and Saupe (1952) range from .42 to .89 for a single observer's measurement, and from .59 to .94 for a team's measurement. In each case, the Competency dimension was least reliable, while Social Organization was most reliable. The single observer coefficients are, according to the authors, "similar in meaning to the correlations of 2 forms of a test (Cornell, Lindvall, and Saupe, 1952, p. 36)." The team estimates, however, are best described as indices of observer agreement (Medley and Mitzel, 1963). As such, the latter assess the objectivity of the Digest, but do not indicate its power to discriminate among classrooms. In a later study, Medley and Mitzel (1958) produced reliability coefficients which do indicate the discriminating power of the dimensions, as well as indices of observer agreement. Their estimates, based on 25-minute observations of 33 New York City classrooms made by 6 observers in teams of 2, range from .00 (Climate—Pupil) to .66 (Social Organization) for interobserver agreement, and from .00 (Climate—Pupil) to .42 (Variety) for discriminating power. The lower reliabilities obtained by Medley and Mitzel are perhaps a reflection of the more homogeneous sample used in their study (Medley and Mitzel, 1963).

Validity: During the development of the Classroom Observation Code Digest, 118 teachers responded to a questionnaire asking them to name the dimensions they considered most likely to distinguish one classroom from another. Responses to this questionnaire served as a check on the validity of the dimensions included on the Digest. The results were analyzed, categorized, and found to support the authors' original selection of dimensions (Cornell, Lindvall, and Saupe, 1952).

Further evidence of validity was provided in a pilot study of the Digest, results of which indicated that the dimensions significantly discriminate among classrooms. Each dimension was found to differ significantly on at least 1 of 5 sources of variation (Cornell, Lindvall and Saupe, 1952).

Norms: Data used in the pilot studies were collected from 32 classrooms in 4 "rural/urban" school districts in a central Illinois county and 4 "suburban-metropolitan" districts in a high socioeconomic area near Chicago. For each school system, 1 class was observed at each of 4 grade levels: fourth, sixth, eighth, and tenth (Cornell, Lindvall, and Saupe, 1952).

Data have also been collected in 33 elementary classrooms in New York City (Medley and Mitzel, 1958).

Procedures for Use: The observer uses the Digest as a guide in recording behavior on the Observation Schedule, or tally sheet. For the first 4 dimensions, he or she enters a number, representing the appropriate description of the classroom situation, for each 5-minute interval in a 1-hour visit. For the fifth dimension, each kind of behavior observed during 5-minute intervals is checked. Behaviors observed on the remaining dimensions are recorded only once during each visit.

The scoring procedure yields a single value for each dimension except Variety, which is scored once for the number of different activities observed, and once for the duration of these activities.

Comments: The Classroom Observation Code Digest is an economical and relatively uncomplicated measure of classroom characteristics. Observers need only about 15 hours of training to code a wide variety of items, many of which require little inference.

References: Cornell, F. G., C. M. Lindvall, and J. L. Saupe, *An exploratory measurement of individualities of schools and classrooms.* Urbana, Illinois: Bureau of Educational Research, College of Education, University of Illinois, 1952.

Medley, D. M. and H. E. Mitzel, "Application of analysis of variance to the estimation of the reliability of observations of teachers' classroom behavior." *Journal of Experimental Education*, 1958, **27**, 23–25.

Medley, D. M. and H. E. Mitzel, " Measuring classroom behavior by systematic observation," in N. L. Gage (ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Classroom Observation Record (Cl. Obs. Record #4.)

AUTHOR: Henry Morrison McGee
AUTHOR/PUBLISHER ADDRESS: Source in McGee (1955)

CRITERIA EVALUATED: 3, 7, 13, 16,

SUBJECT OF OBSERVATION

__x__ TEACHER AND PUPILS
_____ TEACHER ONLY
_____ PUPIL ONLY
_____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
_____ 1 ONLY
__x__ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

__x__ LIVE
_____ LIVE WITH CODING EQUIPMENT
_____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
__x__ AFFECTIVE
_____ COGNITIVE
__x__ PROCEDURE OR ROUTINE
_____ ACTIVITY
__x__ SOCIOLOGICAL STRUCTURE
_____ OTHER

SETTINGS

__x__ CLASSROOM, ANY SUBJECT
_____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

__x__ CATEGORY CHANGE
_____ TOPIC CHANGE

_____ SPEAKER CHANGE
_____ TIME UNIT
_____ TIME SAMPLE
_____ RATING SCALE
_____ OTHER

OBSERVER(S) REQUIRED

__x__ 1 PERSON
_____ TEAM OF 2 PERSONS
_____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
__x__ RESEARCH
_____ TRAINING
_____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 441-443.

Name: CLASSROOM OBSERVATION RECORD

Author: Henry Morrison McGee

Type of Measure: Observation coding system

Availability: McGee, H. M., "Measurement of authoritarianism and its relation to teachers' classroom behavior." *Genetic Psychology Monographs*, 1955, **52**, 89-146.

Description: The Classroom Observation Record was developed by McGee (1955) to assess classroom behavior related to authoritarianism in teachers. The initial form of the Record contained 13 categories of teacher behavior and 4 categories of pupil behavior, all presumed on the basis of theoretical and empirical studies, notably those reported in *The Authoritarian Personality* (Adorno, Frenkel-Brunswick, Levinson, and Sanford, 1950), to be associated with authoritarianism. Each of these categories was accompanied by a set of specific behaviors defining it, and a hypothesis concerning its connection with authoritarianism. The behaviors composing all 13 categories in combination formed a Glossary of terms to be used by the observer in coding classroom activity.

A preliminary study of 50 elementary and secondary school teachers revealed a number of deficiencies in both the Record and the Glossary. As a result, descriptive behaviors with low discriminating power were eliminated, and several categories were revised to accommodate the dichotomous nature of the authoritarian character structure.

The dimensions of teacher behavior included on the revised form of the Record are: Aloof vs. Approachable; Unresponsive vs. Responsive; Dominative vs. Integrative; Irresponsible vs. Responsible; Harsh vs. Kindly; Immature vs. Mature; Inflexible vs. Adaptable; Insensitive vs. Sensitive; Narrow vs. Broad; Partial vs. Fair; Pessimistic vs. Optimistic; Stereotyped vs. Liberal; and Suspicious vs. Trusting. Pupil behavior categories are: Unresponsive vs. Responsive; Dependent vs. Initiating; Obstructive vs. Responsible; and Uncertain vs. Confident.

Sample Items: Each specific descriptive behavior included in the Record's 17 categories relates to 1 or more of the variables comprising the authoritarian syndrome. Furthermore, each is observable. The following behaviors are similar to those which are coded on the Classroom Observation Record.

Variable	*Behavior*
Harsh vs. Kindly	1. Teacher ridiculed child in front of class. 2. Teacher pointed out positive aspects of pupil's performance.

Stereotyped vs. Liberal	1. Teacher treated pupils as a homogeneous group, rather than as individuals. 2. Teacher explored problems, leaving room for questions and exceptions.

Reliability: Reliability of the obtained behavior scores was computed on a sample of 150 teachers, classified according to volunteer/nonvolunteer status, sex, teaching level, and minority/majority group membership. Pearson *rs*, corrected by the Spearman-Brown formula, ranged from .85 to .97. Overall reliability was estimated at .90 (McGee, 1955).

Agreement between observations was computed by correlating the first and second observations of 2 coders, yielding a coefficient of .74. Also, one of these coders and a timed observer simultaneously but independently recorded the behavior of 10% of the sample ($N = 15$). A rank-difference correlation of .89 was obtained between the 2 sets of scores (McGee, 1955).

Validity: Evidence for validity is drawn from 3 sources: item analysis; relationship to similar constructs; and differences between groups.

Internal consistency of the behavioral dimensions, established through item analysis, suggests validity by demonstrating that even the least discriminating categories significantly differentiated between high and low scorers (McGee, 1950).

Empirical validity is indicated by a correlation of .58 between mean behavior scores obtained on the Record and a modified form of the F-Scale, a measure of authoritarianism devised by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1955).

Further evidence of the validity of teacher scores is provided by a highly significant critical ratio of 3.3 between the mean pupil behavior score obtained in classrooms of less authoritarian teachers and that obtained in classrooms of more authoritarian teachers (McGee, 1955).

Norms: Normative data are available for the 184 regular classroom teachers whose behavior was originally coded on the Classroom Observation Record. This sample included men and women, 25 to 30 years of age, teaching at the elementary or secondary level in the public schools of Oakland, California. No one in the sample had taught more than 3 years (McGee, 1955).

Procedures for Use: Before employing the Classroom Observation Record, the user should carefully train prospective observers and then pilot-test the system for agreement among them. This procedure may take as long as 4 weeks.

When observers are ready to enter the classroom, they should do so unobtrusively, situating themselves at the side or rear of the room. Behaviors are recorded as they are observed, unless coding appears to be causing some anxiety on the subjects' part, in which case observed behaviors may be entered immediately after class. An observation period of 45 to 50 minutes is considered optimal length.

Numerical values are assigned to each category according to the number of specific behaviors recorded in that category during a single observation period. For the 13 categories of teacher behavior, total scores can range from 13 to 91 points, a high score indicating strong authoritarian behavior.

Comments: While this instrument does not claim to represent all features of the authoritarian personality, it does include a good sample of those which are commonly expressed in the classroom. Although it requires extensive observer training, the Classroom Observation Record can be a valuable research tool.

References: Adorno, T. W., E. Frenkel-Brunswik, D. S. Levinson, and R. N. Sanford, *The authoritarian personality.* New York: Harper, 1950.

McGee, H. M., "Measurement of authoritarianism and its relation to teachers' classroom behavior." *Genetic Psychology Monographs,* 1955, **52**, 89-146.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Classroom Creativity Observation Schedule (Classroom Creativity #5)

AUTHOR: David A. Denny, Reuben Rusch, Sammie Ives
AUTHOR/PUBLISHER ADDRESS: D. A. Denny, State University College

CRITERIA EVALUATED: 10, 13, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☒ TRAINING
☒ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 219-222.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Classroom Observation Scales (Classroom Obs. Scales #6)

AUTHOR: Edmund T. Emmer
AUTHOR/PUBLISHER ADDRESS: Research and Development Center for Teacher Education

CRITERIA EVALUATED: 9, 10, 11, 14, 15, 16, 17, 19

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

___ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
x PROCEDURE OR ROUTINE
x ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
___ TOPIC CHANGE
___ SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
___ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
x TRAINING
x EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 444-446.

Name: CLASSROOM OBSERVATION SCALES

Author: Edmund T. Emmer

Type of Measure: Observation coding system

Availability: Emmer, E. T. *Classroom observation scales.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1971.

Description: The Classroom Observation Scales (COS) are intended to measure "classroom process variables for research on teaching and to provide an instructional tool in teacher education (Emmer, 1971, p. i)." They are designed for use in the classroom by trained observers. The 12 variables measured by this instrument are listed and briefly described below.

1. *Pupil Attention* refers to student orientation toward the teacher, the task at hand, or the predominant classroom activity.
2. *Teacher-Initiated Problem Solving* indicates a mode of teacher behavior characterized by frequent presentation of questions and problems to the entire class.
3. *Pupil-to-Pupil Interaction* includes substantive utterances of one student directed at another or at a group of students, as well as indirect or unsolicited responses to the teacher.
4. *Teacher Presentation* refers to the relative amount of class time devoted to teacher presentation of substantive information.
5. *Negative Affect* is used to record behaviors which reflect negative or hostile feelings on the part of either the teacher or the pupils.
6. *Positive Affect* encompasses teacher behaviors which support or show positive feeling for pupils and their behavior.
7. *Higher Level Cognitive Student Behavior* refers to pupil verbalizations which reflect more complex cognitive processes.
8. *Passive Pupil Behavior* measures the extent to which pupils demonstrate passive behaviors, such as visual wandering or withdrawal from engagement with surroundings.
9. *Convergent-Evaluative Interaction* is characterized by a focus upon obtaining the correct answer to a teacher question, with little or no attempt to continue contact once the answer has been obtained.
10. *Task Orientation* measures the degree to which the teacher works toward content-related, substantive goals.
11. *Clarity* refers to the degree to which the teacher's presentations and interactions are understood by the pupils.
12. *Enthusiasm* indicates the extent to which the teacher demonstrates interest, vitality, and involvement in his subject and his teaching.

Sample Items: The COS manual (Emmer, 1971) lists specific behaviors and interactions typical of those which are recorded on each of the 12 scales. For example, attention (Scale 1) may be inferred from the following pupil behaviors: maintaining eye contact with the teacher; working on an assigned activity; or raising hand to volunteer a response.

Reliability: Reliability of the COS was established using data collected by graduate and undergraduate students who received approximately 6 hours of training, including practice observations of both videotapes and live classrooms. For each of 2 data sets, the manual presents 2 indices of interobserver agreement (intraclass correlations and average difference scores) along with estimates of stability. Intraclass correlations for data set 1 yielded reliability estimates for single observer scores ranging from .07 (Task Orientation) to .71 (Teacher Presentation). Corresponding correlations for data set 2 ranged from .06 (Higher Level Cognitive Pupil Behavior) to .81 (Pupil Attention). Reliability estimates for the average of 2 observers' scores ranged from .13 (Task Orientation) to .83 (Teacher Orientation) for data set 1, and from .12 (Higher Level Cognitive Pupil Behavior) to .89 (Pupil Attention) for data set 2. Average difference scores ranged from .31 (Pupil Attention) to .90 (Pupil-to-Pupil Interaction) for data set 1, and from .59 (Pupil Attention, Teacher Presentation, and Enthusiasm) to .90 (Teacher-Initiated Problem Solving) for data set 2.

Stability estimates were computed for single observations and for the average of 5 observations of each class. The former ranged from .08 (Task Orientation) to .65 (Enthusiasm) for data set 1 and from .10 (Convergent-Evaluative Interaction) to .64 (Pupil Attention) for data set 2. Stability estimates for an average of 5 observations ranged from .27 (Task Orientation) to .89 (Enthusiasm) for data set 1, and from .30 (Convergent-Evaluative Interaction) to .87 (Pupil Attention) for data set 2.

Validity: All but 3 of the 12 Classroom Observation Scales are empirically based. In developing the scales, observations were collected on 5 occasions in 28 fifth- and eighth-grade classrooms, using 5 coding systems: CASES (Spaulding, 1970); CCS (Emmer and Albrecht, 1970); Dyadic Interaction System (Good and Brophy, 1970); FAIR (Fuller, 1970); and OScAR 5V (Medley et al., 1968). Categories in each of these systems were intercorrelated, and these intercorrelations were then factor analyzed. The factors which emerged were intercorrelated across systems to yield common factors. Scales were defined using the behavioral categories associated with each factor.

Three scales (Task Orientation, Clarity, and Enthusiasm) not derived through factor analysis were also included since they had been associated with positive pupil outcomes in several different studies (Emmer, 1971).

Norms: The Classroom Observation Scales have been used in several research studies. The samples which served as the source for reliability data were composed of 15 fifth-grade classrooms and 31 first- and second-grade class-

rooms, each videotaped on 4 or 5 occasions and coded independently by 2 observers.

Procedures for Use: In using the COS, the observer should be situated so that he can see both the teacher and the pupils. Ideally, he should be able to observe the faces of all students. Thus situated, he records, at 15-minute intervals, observed behaviors on a printed coding form.

Comments: The Classroom Observation Scales are easy to learn and use, and they are, for the most part, behaviorally based. They may be effectively used in research, evaluation, or teacher performance feedback. However, the user should note that several of the scales (Higher Level Cognitive Pupil Behavior, Task Orientation, and Passive Pupil Behavior) have produced marginal or inconsistent reliability data. Furthermore, in using the scales, at least 4 or 5 observations should be collected. Otherwise, the user may obtain data which are not sufficiently reliable.

References: Brophy, J. E., and T. L. Good, *Teacher-child dyadic interaction: A manual for coding classroom behavior.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

Emmer, E. T., *Classroom observation scales.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1971.

Emmer, E. T. and I. D. Albrecht, *The Cognitive Component System (CCS) manual.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

Fuller, F. F., *FAIR system manual.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

Medley, D. M., C. G. Schluck, and N. P. Ames, *Assessing the learning environment in the classroom: A manual for users of OScAR 5V.* Princeton, New Jersey: Educational Testing Service, 1968.

Spaulding, R. L., *Classroom behavior analysis and treatment.* San Jose, California: San Jose State College, 1970.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Cognitive Component System (Cognitive Components #7)

AUTHOR: E. T. Emmer, I. D. Albrecht
AUTHOR/PUBLISHER ADDRESS: Research and Development Center for Teacher Education, University of Texas, Austin, Texas

CRITERIA EVALUATED: 3, 13, 19

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☐ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☒ TOPIC CHANGE
☐ SPEAKER CHANGE
☒ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 447-448.

Name: COGNITIVE COMPONENT SYSTEM

Authors: Edmund T. Emmer and I. Dorothy Albrecht

Type of Measure: Observation coding system

Availability: Emmer, E. T., and I. D. Albrecht, *The Cognitive Component System (CCS) manual.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

Description: The Cognitive Component System (CCS) was designed to measure the cognitive functioning of students and teachers in the classroom. It is used by trained observers, who code cognitive behaviors according to 5 categories: Association, Description, Conceptualization, Generalization, and Inference. These categories are tallied in a manner which distinguishes teacher and student behaviors as well as presenting, responding, and soliciting behaviors. These distinctions produce 30 classifications of cognitive behavior, which are supplemented by 3 activity categories—Directing and Managing, Individual Seat Work, and Group Work—bringing the total number of CCS discriminations to 33.

Sample Items: The CCS manual explicitly defines the system's categories and lists behaviors which are typically coded under each heading. For example, Association is defined as "verbatim reproduction of verbal sequences, a stimulus/response chain without apparent mediators (Emmer and Albrecht, p. 5)." Teacher questions and student responses involving memorization, drill, and recall are coded in this category.

Reliability: Estimates of between-observer agreement, based on intraclass correlations, are presented in the manual for each CCS category. These range from .11 (Student Responds Inference) to .98 (Teacher Solicits Association), with most being in the .80s or .90s. Reliability coefficients are not reported for 7 categories in which insufficient behavior was observed.

Estimates of between-observer agreement for combined CCS categories range from .83 to .98.

Validity: The development of the CCS and the selection of its behavioral categories are not described in the manual. The system's theoretical rationale is not discussed, and no evidence from which to infer validity is provided.

Norms: Data in the manual are based on 15 $\frac{1}{2}$-hour videotapes made in fifth- and eighth-grade classrooms, with 2 observers coding each tape.

Procedures for Use: CCS codings can be made from audiotapes, videotapes, or live observations of classroom interaction. The observer indicates each cognitive statement on a tally sheet. Each entry on the sheet describes 1 unit of behavior—usually a sentence—which represents a single, complete thought. If a thought is repeated (if, for example, a teacher restates a student remark), it is coded only when it first occurs, unless it has been modified or expanded in some way.

Noncognitive behaviors, such as affective communications or routine response acknowledgements, are ignored. This rule, however, does not apply to the 3 classroom activity categories, which are included to provide an estimate of the amount of seat work, group work, and nonsubstantive activity (directing and managing). Behaviors falling in these 3 categories are coded in time, rather than thought, units. The observer records these behaviors at 3-second intervals—unless they are occurring over an extended period, in which case he or she merely "times" them with a stopwatch and enters the duration of occurrence directly onto the tally sheet.

After the observation is completed, entries on the tally sheet are totalled. These frequency counts can be used individually or in combination to yield a variety of information.

Comments: The Cognitive Component System is easy to learn and to use. Coders can be trained in a relatively brief period of time, and, since tallying is done using paper and pencil only, no elaborate or expensive equipment is required. Although CCS reliabilities are generally high, this system could be more confidently recommended if its manual provided some information from which to infer validity.

Reference: Emmer, E. T., and I. D. Albrecht, *The Cognitive Component System (CCS) manual.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Connors-Eisenberg System (Connors-Eisenberg #8)

AUTHOR: C. Keith Connors, Leon Eisenberg
AUTHOR/PUBLISHER ADDRESS:

CRITERIA EVALUATED: 1, 11, 13, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 153-155.

Name: CONNERS-EISENBERG OBSERVATION SYSTEM

Authors: C. Keith Conners and Leon Eisenberg

Type of Measure: Observation coding system

Availability: Conners, C. K., and L. Eisenberg, *The effect of teacher behavior on verbal intelligence in Operation Headstart children.* Baltimore, Maryland: School of Medicine, Johns Hopkins University, 1966. (ERIC Document Reproduction Service No. 010 782)

Description: This observation system is designed to measure 3 sets of teacher and classroom variables. First, observers classify teacher behavior *episodes* according to Communication, Management, and Encouragement. Second, they record the degree to which teacher *activities* promote the following "values":

1. Development of an adequate self-concept, emotional stability, and a sense of security;
2. Intellectual growth;
3. Personal responsibility for private or community property;
4. Cultural habit training;
5. Consideration for others;
6. Achievement orientation;
7. Development of physical abilities and skills;
8. Creativity; and
9. Obedience and self-control.

Third, observers make *overall judgments* of teacher performance on four 6-point scales: Warmth vs. Coldness; Permissiveness vs. Restrictiveness; Active vs. Passive; and Variety vs. Non-Variety.

Sample Items: Observations of teacher behavior are recorded as discrete episodes, as suggested by Baldwin (1960) and Reichenberg-Hackett (1962). An episode is defined as "a change in the triangular relationship between teacher, child, and environment, no matter how small this change might be (Conners and Eisenberg, 1966, p. 4)." A change of subject, for example, or a shift of the teacher's attention, signals the end of one episode and the beginning of another. Each statement made by the teacher is recorded verbatim.

Related episodes are grouped to form teacher activities, which are then scored according to the values listed in the previous section.

Reliability: To determine interobserver agreement, 2 raters independently scored sample protocols for all categories. Spearman rank-order correlation coefficients ranged from .79 (promotion of obedience and self-control) to .97 (promotion of physical abilities and skills), with a mean of .87.

Data collected by all 8 observers on 6 teachers indicated that the greatest variation from one observer to another occurred in the overall teacher ratings. Since the 4 global ratings were highly intercorrelated, they were combined to

form a single "good vs. bad" score. An analysis of variance was then performed on the data, and an intraclass correlation of .82 was computed. This coefficient, which is equivalent to the average intercorrelation among all 8 observers, may be considered a measure of the reliability of individual scores (Conners and Eisenberg, 1966).

No data on the reliability of episode definition were collected. However, according to Conners and Eisenberg (1966), training sessions conducted prior to the observations indicated that raters generally agreed in determining episode boundaries.

Validity: Validity is suggested by correlations between teacher characteristics, as recorded on the Conners-Eisenberg system, and the intellectual growth of pupils, as measured by the Peabody Picture Vocabulary Test (PPVT). For example, teacher activities judged as promoting intellectual growth were strongly associated with increased PPVT scores. Similarly, teacher behavior which placed a high value on property rights and care of materials produced little change in PPVT scores. Teachers rated as warm, active, permissive, and varied produced more intellectual growth than those judged low on these traits (Conners and Eisenberg, 1966).

Norms: Subjects in the Conners and Eisenberg study were 38 teachers (32 from public schools and 6 from a privately operated church nursery school) and 500 children participating in a Headstart program in Baltimore City in the summer of 1965.

Procedures for Use: In using this system, observers take general notes on an entire day's activities, but record a 1-hour period in great detail. As unobtrusively as possible, they record teacher behavior as discrete episodes. When the observation interval is over, these episodes are scored in 2 ways. According to the first procedure, each episode is scored for 1 or more of the following variables: (1) Communication (to the child, the group, or an adult); (2) Management (altering a child's activity, either positively, by suggesting a substitute activity, or negatively, by simply interrupting or stopping the activity); and (3) Encouragement. The second scoring procedure groups related, consecutive episodes to form teacher activities, which are then scored according to the degree to which they foster 9 specific "values." A teacher is classified as being "high," "medium," or "low" on each value, relative to other teachers.

Finally, each teacher is given an overall score on Warmth, Permissiveness, Activity, and Variety.

Comments: The Conners-Eisenberg Observation System is appropriate for use in assessing elementary or nursery school teachers. It yields 3 sets of scores from 1 set of data and, therefore, is an economical research tool. Although the categories appear to require considerable inference on the observer's part, indices of interrater agreement are high.

References: Baldwin, A. L., "The study of child behavior and development," in P.

Mussen (ed.), *Handbook of research methods in child development*. New York: Wiley, 1960.

Conners, C. K., and L. Eisenberg, *The effect of teacher behavior on verbal intelligence in Operation Headstart children*. Baltimore, Maryland: School of Medicine, Johns Hopkins University, 1966. (ERIC Document Reproduction Service No. ED 010 782)

Reichenberg-Hackett, W., "Practices, attitudes, and values in nursery group education." *Psychological Reports*, 1962, **10**, 151–172.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Coping Analysis Schedule for Educational Setting (CASES) (CASES #9)

AUTHOR: Robert L. Spaulding
AUTHOR/PUBLISHER ADDRESS: Duke University, Durham, North Carolina

CRITERIA EVALUATED: 9, 15

SUBJECT OF OBSERVATION

TEACHER AND PUPILS
TEACHER ONLY
[✓ PUPIL ONLY]
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
1 ONLY
[✓ MORE THAN 2 PEOPLE IN CLASSROOM]

COLLECTION METHODS

[✓ LIVE]
LIVE WITH CODING EQUIPMENT
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
[✓ AFFECTIVE]
COGNITIVE
PROCEDURE OR ROUTINE
[✓ ACTIVITY]
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

CATEGORY CHANGE
TOPIC CHANGE

SPEAKER CHANGE
TIME UNIT
[✓ TIME SAMPLE]
RATING SCALE
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
[✓ RESEARCH]
[✓ TRAINING]
EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 587-589.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: CERLI Verbal-Behavior Classification System (CERLI Verbal-Beh #10)

AUTHOR: Cooperative Educational Research Laboratory, Inc.
AUTHOR/PUBLISHER ADDRESS: Cerli, North Field, Illinois

CRITERIA EVALUATED: 13, 19

SUBJECT OF OBSERVATION

_X_ TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
_X_ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

___ 1 ONLY
_X_ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

_X_ LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

_X_ AFFECTIVE
_X_ COGNITIVE
_X_ PROCEDURE OR ROUTINE
___ ACTIVITY
_X_ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

_X_ CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

_X_ CATEGORY CHANGE
___ TOPIC CHANGE

_X_ SPEAKER CHANGE
_X_ TIME UNIT
___ TIME SAMPLE
___ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

_X_ 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

_X_ RESEARCH
_X_ TRAINING
___ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 215-218.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Disruptive Behavior Schedule (Disruptive Behaviour #11)

AUTHOR: Wesley C. Becker, Charles H. Madsen, Jr., Carole R. Arnold, Don R. Thomas
AUTHOR/PUBLISHER ADDRESS:

CRITERIA EVALUATED: 9

SUBJECT OF OBSERVATION

____ TEACHER AND PUPILS
____ TEACHER ONLY
x PUPIL ONLY
____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
____ LIVE WITH CODING EQUIPMENT
____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

____ AFFECTIVE
____ COGNITIVE
____ PROCEDURE OR ROUTINE
x ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
____ TOPIC CHANGE

____ SPEAKER CHANGE
x TIME UNIT
____ TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
____ TRAINING
____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 351-353.

Name: DISRUPTIVE BEHAVIOR SCHEDULE

Authors: Wesley C. Becker, Charles H. Madsen, Jr., Carole R. Arnold, and Don R. Thomas

Type of Measure: Observation coding system

Availability: Becker, W. C., C. H. Madsen, Jr., C. R. Arnold, and D. R. Thomas, "The contingent use of teacher attention and praise in reducing classroom behavior problems." *Journal of Special Education*, 1967, **1**(3), 287-307.

Description: The Disruptive Behavior Schedule was developed in a study of behavior modification in the elementary school classroom (Becker, Madsen, Arnold, and Thomas, 1967). After "problem" children had been selected for observation, preliminary coding categories were constructed to reflect pupil behaviors which: (1) interfered with classroom learning; (2) violated classroom rules; (3) were considered undesirable by the teacher (e.g., thumbsucking); (4) were similar within categories; (5) were different between categories; and (6) were observable rather than inferred. These categories were tested and revised over a 4-week period to insure that the pupil behaviors represented did in fact occur frequently enough to constitute a problem and be reliably rated. When interrater agreement of at least 80% could not be attained, the category in question was abandoned or redefined. These selection procedures produced the coding categories listed below, 7 of which were applied to each child observed and 4 of which were devised especially for particular children. The final category was included to record each child's "time on task."

A. General Categories of Disruptive Behavior
   1. Gross Motor Behaviors
   2. Disruptive Noise with Objects
   3. Aggression—Disturbing Others Directly
   4. Orienting Responses
   5. Blurting Out, Commenting, and Vocal Noise
   6. Talking
   7. Other

B. Special Categories of Disruptive Behavior
   1. Improper Position
   2. Sucking
   3. Bossing
   4. Ignoring

C. Relevant Behavior

Sample Items: Specific behaviors coded under each category are listed by the authors. For example, Orienting Responses includes "turning head or head and body to look at another person, showing objects to another child, attending to another child (Becker et al., 1967, p. 290)."

Reliability: Since the Disruptive Behavior Schedule has been employed in a

number of studies, several estimates of interrater reliability are available. When the instrument was initially used, (Becker et al., 1967), reliabilities above 80% were obtained prior to baseline data collection. Checked periodically thereafter, these rarely fell below 80%. A similar level of interobserver agreement was obtained by O'Leary and Becker (1968), who ran 6 reliability checks, averaging 82%. O'Leary, Becker, Evans, and Saudargas (1969), studying the effects of a token reinforcement program in a public school, conducted numerous reliability checks throughout the various stages of their data collection and obtained interrater agreement figures ranging from 68 to 100%, and averaging well above 80%. Using a modified version of the Disruptive Behavior Schedule, O'Leary, Kaufman, Kass, and Drabman (1970) achieved comparable levels of interobserver reliability. In each of these cases, reliability was determined for each child observed by dividing the number of agreements between 2 observers by the total number of agreements plus disagreements. An agreement was scored if both raters recorded the same behavior within a given interval. A disagreement was scored if one observer recorded the behavior and the other did not.

Validity: Specific evidence of validity is not reported. However, the instrument has in several studies recorded a reduction in disruptive pupil behaviors following the introduction of behavior modification procedures into the classroom (Becker et al., 1967; O'Leary and Becker, 1968; O'Leary et al., 1969; O'Leary et al., 1970; Thomas, Nielsen, Kuypers, and Becker, 1967).

Norms: Normative data are not available.

Procedures for Use: In the initial study by Becker et al. (1967), coders underwent 4 weeks of training, during which time they learned to unobtrusively observe children in the classroom. Working in pairs, they observed target children for 20 minutes a day, 4 days a week. Using stop watches, which they synchronized every 5 minutes, the raters observed for 20 seconds and then took 10 seconds to record the observed behaviors according to a simple code. All data were reported as percentages of the time intervals in which deviant behavior occurred. Because classroom activities varied a great deal from day to day, weekly rather than daily averages were reported.

Comments: The Disruptive Behavior Schedule provides an economical and reliable method of assessing problem behavior in the classroom. With its restricted focus, this instrument offers a limited number of well defined categories and an uncomplicated coding system, which has consistently produced a high percentage of interobserver agreement.

References: Becker, W. C., C. H. Madsen, Jr., C. R. Arnold, and D. R. Thomas, "The contingent use of teacher attention and praise in reducing classroom behavior problems." *Journal of Special Education*, 1967, **1**(3), 287–307.

O'Leary, K. D., and W. C. Becker, "The effects of the intensity of a teacher's reprimands on children's behavior." *Journal of School Psychology*, 1968, **7**, 8-11.

O'Leary, K. D., W. C. Becker, M. B. Evans, and R. A. Saudargas, "A token reinforcement program in a public school: A replication and systematic analysis." *Journal of Applied Behavior Analysis*, 1962, **2**, 3–13.

O'Leary, K. D., K. F. Kaufman, R. E. Kass, and R. S. Drabman, "The effects of loud and soft reprimands on the behavior of disruptive students." *Exceptional Children*, 1970, **37**(2), 145–155.

Thomas, D. A., L. J. Nielsen, D. S. Kuypers, and W. C. Becker, "Social reinforcement and remedial instruction in the elimination of a classroom behavior problem." *Journal of Special Education*, 1967, **2**(3), 291–302.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Dodl System (Dodl #12)

AUTHOR: Norman R. Dodl
AUTHOR/PUBLISHER ADDRESS: Florida State University, Tallahassee, Florida

CRITERIA EVALUATED: 13, 16, 17, 19

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☒ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 229-232.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Scale for Rating Effective Teacher Behavior (Effective Teacher #13)

AUTHOR: Dwight E. Beecher
AUTHOR/PUBLISHER ADDRESS: Source in Beecher (1949)

CRITERIA EVALUATED: 7, 10, 17

SUBJECT OF OBSERVATION

[✓ TEACHER AND PUPILS]
TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
1 ONLY
[✓ MORE THAN 2 PEOPLE IN CLASSROOM]

COLLECTION METHODS

[✓ LIVE]
LIVE WITH CODING EQUIPMENT
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
[✓ AFFECTIVE]
[✓ COGNITIVE]
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

[✓ CATEGORY CHANGE]
TOPIC CHANGE

SPEAKER CHANGE
TIME UNIT
TIME SAMPLE
[✓ RATING SCALE]
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
RESEARCH
[✓ TRAINING]
[✓ EVALUATION]

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 165-166.

Name: SCALE FOR RATING EFFECTIVE TEACHER BEHAVIOR

Author: Dwight E. Beecher

Type of Measure: Observer rating form

Availability: Beecher, D. E., *The evaluation of teaching: Backgrounds and concepts.* New York: Syracuse University Press, 1949.

Description: The Scale for Rating Effective Teacher Behavior (SRETB) is a 38-item device intended to measure vital aspects of the student/teacher relationship. Scale development began with a review of studies concerning pupil evaluation of teachers. This review produced 6 teacher rating categories: Fairness; Cheerfulness; Sympathetic Understanding; Control; Ability to Stimulate Response; and Knowledge. Under these categories were classified the specific, observable teacher behaviors which, according to the author, most frequently provoke a reaction among high school students. These behaviors are the scale's items. The observer, after rating the teacher on each of these items, may clarify his ratings with anecdotal comments, space for which is provided on the record sheet.

After several preliminary administrations, Beecher decided to leave the scale in its original form. He suggests (1949) that it be used by teachers for self-evaluation and by student teachers for supplemental training information.

Sample Items: Each item on the SRETB describes an observable teacher behavior. If the behavior described is observed, the item is marked with a plus sign; if its opposite is observed, the item is marked with a minus sign. From 5 to 8 items of the following type appear in each category:

Category	*Item*
Fairness	1. Shows no favorites.
Cheerfulness	2. Is not crabby or cross. Does not nag.
Sympathetic Understanding	3. Always tries to show approval of effort. Encouraging.
Control	4. Maintains good pupil conduct without reprimanding frequently.
Ability to Stimulate Response	5. Elicits responses, enthusiasm from students. Many pupils contribute.
Knowledge	6. Knows subject area well. Can answer students' questions.

Reliability: Beecher (1949) reports 2 reliability coefficients. Interrater agreement between 2 judges of the same teacher was estimated at .79. Also, 2 ratings of the same teacher, taken by the same observer on 2 different occasions, correlated .90.

Validity: Because it was developed directly from research findings related to

pupil/teacher relationships and the variables affecting those relationships, the SRETB can claim some content validity. Evidence of concurrent validity is provided by Beecher (1949), who had a panel of "experts" rate 50 teachers as superior, average, or poor. The same 50 teachers were, at the same time, rated on the SRETB by an observer. Scores on the scale were used to divide the teachers into upper, middle, and lower groups. A contingency coefficient for these data was computed and found to be .72. Beecher, using Yule's Formula, corrected this contingency coefficient to estimate the product-moment correlation between the 2 ratings. The corrected value was .88.

Norms: Norms consist of percentile ranks, divided into 9 deciles, which correspond to computed scores. However, information regarding the number, kind, and location of teachers on which these norms are based is not reported.

Administration and Scoring: Beecher (1949) presents only a page of directions for administering the SRETB. Procedures for training observers are not mentioned, although it seems reasonable to assume that such training would be necessary. Because it takes a good deal of time to rate a single teacher on all 38 items, Beecher suggests that only 25 items be used at a time, thus requiring 2 classroom visits for a complete observation.

Scoring the SRETB involves computing the percentage of items marked positively. Thus, if 38 items are rated, and 34 of these ratings are positive, the teacher's score is $\frac{34}{38} \times 100$, or 89.

Comments: The Scale for Rating Effective Teacher Behavior is a logically developed instrument which can be useful in teacher evaluation or student teacher training. However, its value is limited by poorly defined norms and an absence of instructions for training observers. Furthermore, the "either/or" response format of the SRETB has been questioned (in Buros, 1953) since it allows no room for teacher behavior which falls in between the positive and negative extremes implied in Beecher's items.

References: Beecher, E., *The evaluation of teaching: Backgrounds and concepts.* New York: Syracuse University Press, 1949.

Buros, O. K., *The fourth mental measurements yearbook.* Highland Park, New Jersey: The Gryphon Press, 1953.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Florida Climate and Control System (Florida Climate #14)

AUTHOR: Robert Soar, Ruth Soar, Margorie Ragosta

AUTHOR/PUBLISHER ADDRESS: Institute for Development of Human Resources, College of Education, University of Florida, Gainesville, Florida

CRITERIA EVALUATED: 1, 7, 12, 16, 17

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
____ TEACHER ONLY
____ PUPIL ONLY
____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
____ LIVE WITH CODING EQUIPMENT
____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
____ COGNITIVE
____ PROCEDURE OR ROUTINE
x ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

____ CATEGORY CHANGE
____ TOPIC CHANGE
____ SPEAKER CHANGE
____ TIME UNIT
x TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
____ TRAINING
____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 451-453.

Name: FLORIDA CLIMATE AND CONTROL SYSTEM

Authors: Robert Soar, Ruth Soar, and Marjorie Ragosta

Type of Measure: Observation coding system

Availability: Soar, R., R. Soar, and M. Ragosta, *Florida climate and control system.* Gainesville, Florida: Institute for the Development of Human Resources, College of Education, University of Florida, 1971.

Description: The Florida Climate and Control System (FLACCS) was developed for use in Project Follow Through (Soar, 1973), a study of the educational needs of disadvantaged children. Derived primarily from the South Carolina Observation Record (Soar, 1966), the Hostility-Affection Schedule (Fowler, 1962), and earlier versions of the Observation Schedule and Record (Medley and Mitzel, 1958), FLACCS was designed to record behavioral dimensions not covered by Interaction Analysis. Among these were classroom groupings, the degree of individual versus group work, the physical movement of teacher and pupils, and the nonverbal expression of affect in the classroom.

After several revisions, FLACCS emerged as a 2-part instrument, the first section of which focuses on the teacher's direction and control of the classroom and the pupils' response to such management. This section includes items describing Teacher Activity, Verbal Control, Nonverbal Control, Pupil Activity, Socialization, Materials, and Pupil Interest/Attention. The second portion of FLACCS assesses expression of negative and positive affect under the following categories: Teacher Verbal; Pupil Verbal; Teacher Nonverbal; and Pupil Nonverbal.

The control and affect sections of FLACCS were separately factored by Soar (1973). Items with the highest factor loadings were jointly analyzed, yielding the following 9 factors.

1. Strong Control
2. Pupil Free Choice vs. No Choice
3. Teacher/Pupil Supportive Behavior
4. Nonverbal Gentle Control
5. Gentle Control
6. Work Without the Teacher
7. Pupil Negative Affect
8. Teacher Attention in a Task Setting
9. Teacher Positive Affect

Sample Items: Most of the behavioral items included on FLACCS are specific, demanding little inference on the observer's part (e.g., "smiles, laughs, nods" under Positive Affect, Teacher Nonverbal). Other items, however, are more general, requiring some interpretation by the observer (e.g., "makes disparaging remark" under Negative Affect, Pupil Verbal).

Reliability: Soar (1973) reports correlations between observations taken at 3 points in time for all 9 FLACCS factors. The first and second observations correlated from −.16 to .77, the first and third from .09 to .89 and the second and third from −.05 to .72, indicating considerable variability in factor stability. Soar reports less variability among observers, presenting indices of interobserver agreement calculated by Hoyt's (1955) analysis of variance formula 5. These ranged from .77 to .99 and averaged .92.

Validity: Intercorrelations among the FLACCS factors, computed by Soar (1973), were generally low, indicating some construct validity. Only 2 were above .50—that between Strong Control and Pupil Negative Affect and that between Pupil Free Choice and Teacher Positive Affect. Further evidence of construct validity is provided by group differences which emerged on the FLACCS. Five of the 9 FLACCS factors showed F ratios which significantly discriminated grade levels. In general, kindergarten differed most often from the other grades, as might be expected. Kindergarten classrooms were characterized on the FLACCS by more Pupil Free Choice and Teacher Gentle Control and less Strong Control, Work Without Teacher, and Teacher Attention in a Task Setting.

Evidence of criterion-related validity is provided by correlations between FLACCS factors and other measures of classroom interaction. For example, the FLACCS factor, Pupil Free Choice, related negatively to Convergent Thinking, as measured on the Teacher Practices Observation Record (TPOR), but positively to the TPOR variables, Pupil Free Choice and Exploration of Ideas. This same FLACCS factor, Pupil Free Choice, also correlated positively with Varied Pupil-Initiated Interaction, measured on the Reciprocal Category System (RCS), and negatively with Drill, also on the RCS. Both of these correlations exceeded .50.

Relationships between FLACCS factors and measures of pupil behavior also have bearing on the system's validity. Pupil absences in the upper grade levels correlated positively with Strong Control and negatively with Teacher Positive Affect. In the lower grades, pupil skills gain related positively to Teacher Attention in a Task Setting and negatively to Pupil Free Choice. In general, the expected relationship between teacher expressions of affect and pupil gain did not appear. Instead, pupil gain was associated with the FLACCS dimensions reflecting structuring of classroom activity and pupil involvement with the teacher (Soar, 1973).

Norms: Soar (1973) presents factor means and standard deviations for several elementary grade levels.

Procedures for Use: FLACCS categories are presented on 2 pages, the first reflecting teacher control and pupil reaction, the second, teacher and pupil expression of positive and negative affect. Behaviors are recorded separately on each page. Certain items (such as "Praise") appear on both pages of the instru-

ment. These may be coded once or twice, depending on the use and intent of the item. If a teacher praises a child's work in order to reinforce and encourage him, the item is coded in both the structural and affective areas. If, however, the teacher compliments a child's appearance, a check is placed only on the affective page.

In using the FLACCS, an observer views the classroom for a specified period of time, usually 2 to 5 minutes, and then tallies both pages of the code sheet. Behaviors are recorded only once, even if they recur several times within the brief observation period. This procedure is then repeated. The ratings thus reflect overall impression of the classroom.

Comments: The Florida Climate and Control System offers comprehensive assessment of classroom interaction, requiring little inference on the observer's part. The FLACCS has a strong affective component and is most appropriate for the elementary grades. The somewhat low reliabilities between observations reflect changes in pupil and teacher behaviors over the year and, therefore, should not necessarily be interpreted as a characteristic of the instrument.

References: Fowler, B. D., *Relation of teacher personality characteristics and attitudes to teacher-pupil rapport and emotional climate in the elementary classroom.* Unpublished doctoral dissertation, University of South Carolina, 1962.

Hoyt, C. J., "Relations of certain correlational to variance ratio estimates of test reliability." *Twelfth Yearbook of the National Council of Measurement Used in Education*, 1955, 50–55.

Medley, D. M., and H. E. Mitzel, "A technique for measuring classroom behavior." *Journal of Educational Psychology*, 1958, **49**, 86–92.

Soar, R. S., *An integrative approach to classroom learning.* NIMH project 5-R11 MH 01096. Columbia, South Carolina: University of South Carolina, 1966. (ERIC Document Reproduction Service No. ED 033 749).

Soar, R. S., Final report. *Follow Through classroom process measurement and pupil growth (1970–71).* Gainesville, Florida: Institute for Development of Human Resources, College of Education, University of Florida, 1973.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Fuller Affective Interaction Records 33 (FAIR 33 #15)

AUTHOR: Frances F. Fuller
AUTHOR/PUBLISHER ADDRESS: Research and Development Center for Teacher Education, University of Texas, Austin, Texas

CRITERIA EVALUATED: 7, 14, 17, 19

SUBJECT OF OBSERVATION

- _x_ TEACHER AND PUPILS
- ___ TEACHER ONLY
- ___ PUPIL ONLY
- ___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

- ___ 1 ONLY
- _x_ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

- ___ LIVE
- _x_ LIVE WITH CODING EQUIPMENT
- ___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

- _x_ AFFECTIVE
- ___ COGNITIVE
- ___ PROCEDURE OR ROUTINE
- ___ ACTIVITY
- ___ SOCIOLOGICAL STRUCTURE
- ___ OTHER

SETTINGS

- _x_ CLASSROOM, ANY SUBJECT
- ___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

- _x_ CATEGORY CHANGE
- ___ TOPIC CHANGE
- ___ SPEAKER CHANGE
- ___ TIME UNIT
- ___ TIME SAMPLE
- ___ RATING SCALE
- ___ OTHER

OBSERVER(S) REQUIRED

- _x_ 1 PERSON
- ___ TEAM OF 2 PERSONS
- ___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

- ___ RESEARCH
- _x_ TRAINING
- _x_ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 454-456.

Name: FULLER AFFECTIVE INTERACTION RECORDS 33

Author: Frances F. Fuller

Type of Measure: Observation coding system

Availability: Fuller, F. F. *FAIR system manual. Fuller Affective Interaction Records.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1969.

Description: The FAIR observation system employs trained coders to assess interpersonal behaviors of preservice teachers and their students in self-contained (single classroom) settings. FAIR 33 is based on 5 interpersonal dimensions. The first of these, *Responsiveness*, includes behaviors which occur in response to another's actions (*R*) or serve to initiate a new action (*I*). The second dimension, *Self-Other*, describes actions directed toward the self (*S*) or toward others (*O*). The third dimension, *Approval*, includes behaviors ranging from approval (*A*), to noncommittal (*N*), to disapproval (*D*), and the fourth, *Inclusion*, describes behaviors ranging from an invitation to respond (*I*) to exclusion of another (*E*). The final dimension, *Control*, is comprised of behaviors which permit (*P*) or restrict (*R*) the responses of others. Combinations of the behaviors within these 5 dimensions produce 14 teacher categories, 14 student categories, and 5 noninteractive categories (Fuller, 1969; Manning, 1972).

Sample Items: The manual presents a series of dialogues illustrating the 33 categories generated from the behaviors included in the 5 dimensions. Sample categories produced from dimension combinations appear below (Fuller, 1969, p. 20):

Dimension Combination	*Teacher Category*	*Student Category*
ROAIP (*R*esponds, *O*ther, *A*pproval, *I*nvites, *P*ermits)	*Values.* Values feelings; identifies; shares. Listens attentively. Unqualified acceptance. Includes laughing or being sad *with* someone, "I feel that way too." (Person oriented)	*Zeal.* Student is responsive, participating. Listens attentively. Values or recognizes another's feelings. Waves hand, shows pleasure, appreciation, good mood, laughing *with* someone. "Oh, Boy!" (Affect)

ROAIR (*R*esponds, *O*ther, *A*pproval, *I*nvites, *R*estricts)	*Nurtures.* Teacher gives focused encouragement. Guides. Hints. "Come on, Johnny, you know this one." Gives praise, approval to previous behavior. Smiles. Includes recognition of student volunteer, and "Thank you for helping me." (Affect)	*Encourages.* Encourages teacher or another student to go on. Includes "Thank you for helping me." Gives approval; praises. "You got it right." Includes choosing in a game, election panel.

Reliability: The reliability of mean ratings of two coders is presented in the manual for each FAIR category. Coefficients range from .97 to .00, with an average of .56. Average reliability for teacher codes is .70; average reliability for student codes is .42. For 97% of the data collected, average intercoder agreement is .81 (Fuller, 1969).

Validity: No information on the validity of the FAIR 33 is reported in the manual. However, FAIR 33 categories were constructed from classroom behaviors recorded on sound films and videotapes. These included behaviors which were observed, as well as behaviors which were assumed to exist. For example, teachers were observed praising students, but students were not observed praising teachers. Yet the latter behavior probably occurs, and was therefore included on the FAIR 33. All behaviors observed in 360 films and videotapes have successfully been classified within this category system (Fuller, 1969).

Norms: No normative data are presented.

Procedures for Use: Each time behavior changes, a code is recorded on the FAIR 33. A single behavior may last for 1 second or several minutes. The length of the behavior is recorded according to procedures described in the manual. One coder records the behavior of both teacher and students. It is recommended that a coding desk or flexowriter be used in conjunction with the coding system. The flexowriter produces a punched paper tape which can be easily transferred to magnetic tape for computer processing, thereby permitting large quantities of data to be processed without tallying or keypunching.

Comments: The FAIR 33 categories allow the researcher to view behaviors in relation to one another. The FAIR system offers an efficient means of processing large quantities of observation data. However, because of heavy demands placed on the observers, some mechanical and electronic equipment is needed to eliminate tallying and key punching. FAIR reliabilities are generally high, and the instrument appears to be soundly constructed, although validity data are lacking.

References: Fuller, F. F., *FAIR system manual. Fuller Affective Interaction Records.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1969.

Manning, B. A., *Interpersonal interaction: A selective review.* Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1972.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Hough System (Hough #16)

AUTHOR: John B. Hough
AUTHOR/PUBLISHER ADDRESS: Ohio State University, Columbus, Ohio

CRITERIA EVALUATED: 8, 9, 16, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☐ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE
☐ SPEAKER CHANGE
☒ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☒ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 301-304.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Jackson and Lahaderne System (Jackson, Lahaderne #17)

AUTHOR: Phillip W. Jackson and Henriette M. Lahaderne
AUTHOR/PUBLISHER ADDRESS: Source in Jackson and Lahaderne (1967)

CRITERIA EVALUATED: 7, 9, 12, 15, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☒ AFFECTIVE
☐ COGNITIVE
☒ PROCEDURE OR ROUTINE
☒ ACTIVITY
☒ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 457-458.

Name: JACKSON AND LAHADERNE OBSERVATION SYSTEM

Authors: Phillip W. Jackson and Henriette M. Lahaderne

Type of Measure: Observation coding system

Availability: This instrument is described in: Jackson, P. W., and H. M. Lahaderne, "Inequalities of teacher-pupil contacts." *Psychology in the Schools*, 1967, **4**(3), 204–211.

Description: This observation system was developed by Jackson and Lahaderne (1967) in a study of teacher/pupil interaction in elementary school classrooms. Designed to assess the communication between the teacher and individual students, this instrument requires an observer to sit in the classroom and record the following information for each verbal or nonverbal exchange: (1) the identity of the student involved; (2) the initiator of the message (i.e., teacher or student); and (3) the nature of the communication. Any given communication is classified as primarily Instructional, Managerial, or Prohibitory. Instructional messages refer in some way to curriculum content or educational objectives. Managerial messages, on the other hand, involve the interpretation of classroom rules and the definition of acceptable behavior, while Prohibitory remarks are intended to maintain order or punish misbehavior.

Sample Items: Jackson and Lahaderne (1967) illustrate the classification of several nonverbal behaviors, noting that "the ability to categorize messages that are only partially overheard or that do not entail words is particularly important in elementary classrooms where teachers and students are very mobile and where the occasion for certain messages recurs so frequently that their transmission becomes highly stylized and abbreviated (p. 205)." Thus, they explain that when a student approaches the teacher with an open workbook in his hand, the interaction, though inaudible to the observer, may safely be classified as Instructional. Or, when the teacher snaps his fingers and points to a student, it may be assumed that the message is Prohibitory even though no words are spoken.

The classification of verbal behavior is not similarly exemplified. The authors maintain that observers can usually differentiate the 3 types of communication, though Managerial and Prohibitory messages may sometimes be confused.

Reliability: Reliability data are not reported by Jackson and Lahaderne (1967).

Validity: Though no evidence of criterion-related validity is presented, many of the results which Jackson and Lahaderne obtained using this instrument reveal expected differences and similarities in classroom communication. Data from 4 sixth-grade classes in a relatively homogeneous community indicated a rapid rate of interaction and a preponderance of Instructional exchanges in all of the classrooms. The 4 groups were also quite similar in regard to the initiation of

student/teacher interaction. In each room the teacher controlled the flow of communication, initiating about 80 individual exchanges per hour. Initiation rates for students, however, differed from class to class, and among those rooms with high student initiation rates, the focus of student queries differed.

Further differences emerged with regard to pupil sex, confirming the popular belief that boys are the source of most classroom misbehavior. Though only 44.8 to 58.8% of the children in each class were male, 69.6 to 90.1% of the Prohibitory exchanges involved boys rather than girls. Within each room and within each sex group, however, there remained wide individual differences in teacher/pupil interaction, with 1 or 2 students having fewer than 1 exchange per hour with the teacher and others being in contact with the teacher almost constantly.

Norms: The sample used by Jackson and Lahaderne (1967) was composed of 4 sixth-grade classrooms in 2 schools in a predominantly white, working-class community. Two of the classrooms contained 34 pupils each, and 2 others contained 29 each. Two of the 4 teachers were men and 2 were women.

Procedures for Use: In the Jackson and Lahaderne study (1967), observations were tallied in 20-minute units, 36 of which were recorded for each classroom. These observations, distributed throughout the school week, covered all of the activities of each room. The observation tally sheet required that an entry be made for each transmission of information between the teacher and individual students. Teacher remarks directed at more than 1 student were ignored.

Comments: The primary advantage of the Jackson and Lahaderne Observation System is its simplicity. It allows the observer to record both verbal and nonverbal classroom interaction using relatively few, fairly easily distinguished categories. However, it should be kept in mind that such simplicity may also produce a broad, general picture of classroom communication, which may suffer from lack of detail and depth. This possibility, along with the absence of reliability data, limits this instrument somewhat.

Reference: Jackson, P. W., and H. M. Lahaderne, "Inequalities of teacher-pupil contacts." *Psychology in the Schools*, 1967, 4(3), 204–211.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Joyce System (Joyce #18)

AUTHOR: Bruce R. Joyce
AUTHOR/PUBLISHER ADDRESS: Teachers' College, Columbia University, New York, New York

CRITERIA EVALUATED: 3, 8, 13, 14, 17, 18, 19

SUBJECT OF OBSERVATION

___ TEACHER AND PUPILS
_x_ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

_x_ 1 ONLY
___ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

_x_ LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

_x_ AFFECTIVE
_x_ COGNITIVE
_x_ PROCEDURE OR ROUTINE
___ ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

_x_ CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

_x_ CATEGORY CHANGE
___ TOPIC CHANGE
___ SPEAKER CHANGE
_x_ TIME UNIT
___ TIME SAMPLE
___ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

_x_ 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

_x_ RESEARCH
_x_ TRAINING
___ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 337-341.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Kounin System (Kounin #19)

AUTHOR: Jacob S. Kounin
AUTHOR/PUBLISHER ADDRESS: Wayne State University, Detroit, Michigan

CRITERIA EVALUATED: 4, 5, 7, 9

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☐ LIVE
☐ LIVE WITH CODING EQUIPMENT
☒ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☐ AFFECTIVE
☐ COGNITIVE
☒ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☒ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 343-347.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Kounin, Friesen, and Norton Behavior Coding Procedure (Kounin, Friesen #20)

AUTHOR: Jacob S. Kounin, Wallace V. Friesen, A. Evangeline Norton
AUTHOR/PUBLISHER ADDRESS: Source in Kounin, Friesen, and Norton (1966)

CRITERIA EVALUATED: 4, 5, 7, 9, 11, 14, 17

SUBJECT OF OBSERVATION

- x TEACHER AND PUPILS
- ____ TEACHER ONLY
- ____ PUPIL ONLY
- ____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

- ____ 1 ONLY
- x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

- ____ LIVE
- ____ LIVE WITH CODING EQUIPMENT
- x VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

- ____ AFFECTIVE
- ____ COGNITIVE
- ____ PROCEDURE OR ROUTINE
- x ACTIVITY
- ____ SOCIOLOGICAL STRUCTURE
- ____ OTHER

SETTINGS

- x CLASSROOM, ANY SUBJECT
- ____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

- ____ CATEGORY CHANGE
- ____ TOPIC CHANGE
- ____ SPEAKER CHANGE
- x TIME UNIT
- ____ TIME SAMPLE
- x RATING SCALE
- ____ OTHER

OBSERVER(S) REQUIRED

- x 1 PERSON
- x TEAM OF 2 PERSONS
- x 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

- x RESEARCH
- ____ TRAINING
- ____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 459-461.

Name: KOUNIN, FRIESEN, AND NORTON BEHAVIOR CODING PROCEDURE

Authors: Jacob S. Kounin, Wallace V. Friesen, and A. Evangeline Norton

Type of Measure: Observation coding system

Availability: Kounin, J. S., W. V. Friesen, and A. E. Norton, "Managing emotionally disturbed children in regular classrooms." *Journal of Educational Psychology*, 1966, **57**(1), 1–13.

Description: Developed in a study of the relationship between teaching style and the behavior of emotionally disturbed children in regular elementary classrooms (Kounin, Friesen, and Norton, 1966), this observation system focuses on the pupil's academic activity and the teacher's managerial techniques. In developing codes for teacher style, the authors were guided by several criteria. The categories were required to reflect: (1) concrete behaviors or managerial methods, rather than traits or characteristics; (2) dimensions on which teachers differed; and (3) behaviors which were relatively free of commitment to any philosophical orientation or value system. The teacher categories selected are listed below.

I. Desist Techniques
  A. Clarity
  B. Firmness
  C. Child treatment
    1. Negative
    2. Neutral
    3. Positive
  D. Intensity
  E. Focus
    1. Misbehavior
    2. Task
  F. Overlappingness
  G. With-it-ness
  H. Degree of success (coded on a 7-point scale)

II. Transition Incidents
  A. Behavior mode of children at time of transition
  B. Delineation
    1. Simple
    2. With clarity
    3. With group alerting
    4. With challenge/arousal
    5. With zest
  C. Attention property of the signal

D. Presence of anti-resolution properties
E. Appropriateness of timing
F. Children's immediate reaction

The categories of pupil behavior selected for observation apply only to the child's academic activity. These categories, listed below, produce 2 indices of student behavior: Work Involvement rate and Deviancy rate.

I. Definitely and completely involved in work
II. Probably involved in work
III. Definitely not involved in work
IV. Restless
V. Languishing
VI. Languishing and restless
VII. Engaged in task-related deviancy
VIII. Engaged in non-task-related deviancy
  A. Mild
  B Serious
IX. Contagion of deviancy

Sample Items: Kounin et al. (1966) define the categories included on this system and list behaviors typically coded under each. For example, "With-it-ness" refers to the extent to which the teacher communicates to the pupils that he knows what is happening in the classroom. A teacher indicates that he is not "with-it" when he acts upon a deviancy too late, or when he attends to a minor disturbance while a major deviancy is occurring behind his back.

Reliability: For the pupil behavior categories, interrater agreement between various pairs of 4 observers ranged from 82 to 96% and averaged 92%. The teacher behavior category, Degree of Success, produced an average of 74% agreement among observers, while the remaining teacher Desist categories yielded intercoder agreement averaging 90% (Kounin et al., 1966).

Validity: Evidence for validity is based on findings of Kounin et al. (1966), who used this system to observe emotionally disturbed and nondisturbed children in a variety of classroom subsettings. They obtained group differences and category intercorrelations which were almost all in expected directions. For example, in all classroom subsettings, nondisturbed children received higher work involvement and lower deviancy scores than disturbed children. The behavior of all children, however, varied with classroom subsetting, the highest degree of school-appropriate behavior occurring in subgroup recitation periods and the lowest occurring during seatwork situations. In all settings, the amount of school-appropriate behavior manifested by nondisturbed children was positively related to that exhibited by disturbed children.

Correlations between observation categories were, for the most part predicted by the authors. Although Overlappingness (the extent to which the

teacher attends to 2 aspects of the classroom at once) was not significantly associated with Deviancy or Work Involvement for either group of children, it was related to With-it-ness ($r = .50$, $p < .01$). The more with-it the teacher, the more work-involved and less deviant were the pupils. Teacher behavior during transition periods (in which the children were guided from one activity to another) also correlated with the Work Involvement and Deviancy rates of pupils: high antiresolution scores during these periods were associated with a decline in school-appropriate behavior.

Norms: The authors report mean Work Involvement and Deviancy scores for varying numbers of emotionally disturbed and nondisturbed children in regular elementary classrooms. Figures are provided for four classroom subsettings: (1) subgroup recitation; (2) class recitation; (3) seatwork in teacher's sphere; and (4) seatwork outside teacher's sphere. Normative data are not reported for the categories of teacher behavior (Kounin et al., 1966).

Procedures for Use: In the study by Kounin et al. (1966), classroom behavior was videotaped for entire mornings or afternoons. Observers categorized this behavior from tapes, treating each of the 4 subsettings separately and coding only academic activities. Observed behaviors were tallied in 10-second intervals. The mean percentage of intervals containing a tally for a particular behavior constituted the score for that behavior.

Comments: Although published information regarding this observation system is limited, the Kounin, Friesen, and Norton Behavior Coding Procedure appears to be a useful tool for those researchers who prefer to focus on concrete techniques of classroom management rather than on teacher personality characteristics. It has demonstrated adequate interobserver agreement and preliminary evidence of validity.

Reference: Kounin, J. S., W. V. Friesen, and A. E. Norton, "Managing emotionally disturbed children in regular classrooms." *Journal of Educational Psychology*, 1966, **57**(1), 1–13.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: MacDonald-Zaret System (MacDonald-Zaret #21)

AUTHOR: James B. MacDonald, Esther Zaret
AUTHOR/PUBLISHER ADDRESS: J. B. MacDonald, University of Wisconsin, Milwaukee, Wisconsin

CRITERIA EVALUATED: 3, 8, 14, 16, 17, 19

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

LIVE
LIVE WITH CODING EQUIPMENT
x VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

AFFECTIVE
x COGNITIVE
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
TOPIC CHANGE
x SPEAKER CHANGE
TIME UNIT
TIME SAMPLE
RATING SCALE
OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
TRAINING
EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 371-374.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Observation Procedure for Assessing the Microenvironments of Individual Preschool Children (Microenvironments #22)

AUTHOR: John Dopyera

AUTHOR/PUBLISHER ADDRESS: Dopyera (1969) Available from: William J. Meyer, Director, Syracuse University, Early Childhood Education Center, 340 Huntingdon Hall, 150 Marshall St., Syracuse, NY 13210

CRITERIA EVALUATED: 7, 9, 16, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☐ AFFECTIVE
☐ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 467-468.

Name: OBSERVATION PROCEDURE FOR ASSESSING THE MICRO-ENVIRONMENTS OF INDIVIDUAL PRESCHOOL CHILDREN

Author: John Dopyera

Type of Measure: Observation coding system

Availability: Dopyera, J. *Assessing the micro-environments of individual preschool children.* Syracuse, New York: Head Start Evaluation Project, Syracuse University, 1969. This report may be obtained from William J. Meyer, Director, Syracuse University Early Childhood Education Center, 340 Huntington Hall, 150 Marshall Street, Syracuse, New York 13210.

Description: This observation procedure was developed in an evaluation of compensatory preschool programs to assess the micro-environments of children. The term "micro-environment" was used to describe (1) events in the child's immediate environment which were sensorily accessible to him or her and (2) other setting phenomena which served as a medium for the child's behavior. These events and phenomena were recorded in two 2-minute observations, the first of which concentrated on the following aspects of each child's behavior or environment:

1. Physical location and posture;
2. Focus and coordination;
3. Physical environments encountered (i.e., toys, people, equipment);
4. Affective verbal or vocal behavior;
5. Affective verbal or vocal behavior of other children in the immediate vicinity;
6. Interpersonal verbal and nonverbal encounters between the child and any adult or any other child;
7. Conditions which might affect the accuracy of observations;
8. Behavioral indicators of the child.

A second 2-minute observation provided the following information:

1. Context of the child's behavior;
2. The child's activity;
3. The group's activity;
4. Codes for child, peer, teacher, aide, or other adult activity;
5. Minimum distances between the child and others;
6. Sanctions given by adults to the target child or the other children.

Dopyera (1969) reduced data derived from the above categories to 33 variables representing 7 factors, 5 of which were interpretable. These 5 were labeled: Adult Directed and Highly Controlled Small Group Activities; Diffuse Activity

Structure; Undisrupted Independent Effort; Adult Tolerated Nonsettled Behavior; and Subject-Directed Striving.

Sample Items: Dopyera (1969) provides examples of the kinds of behaviors and environmental characteristics coded under each category on the system. "Behavioral indicators," for example, includes exaggerated gait, tics, and other physical idiosyncrasies of the child.

Reliability: Interrater reliabilities were computed for each of the observational items included in the factor analysis. These averaged .56, .59, .59, and .58 for 3 pairs of observers and for all 6 observers combined, respectively.

Validity: Dopyera (1969) related factor scores to a number of criteria. He found that Adult Directed and Highly Controlled Small Group Activities correlated negatively with posttest Stanford-Binet mental age scores. Factor III, Undisrupted Independent Effort, was negatively related to the amount of time the child required to complete the Caldwell Preschool Inventory and the Animal House subtest of the Wechsler Preschool and Primary Scale of Intelligence. Similarly, Factor IV, Adult-Tolerated Nonsettled Behavior, was negatively associated with mental age change. A positive relationship emerged between Factor IV, Subject-Directed Striving, and posttest mental age and Animal House scores.

Norms: The sample used by Dopyera (1969) was drawn from Head Start classes in upstate New York during 1969–70. It included 68 children (males and females, blacks and whites) from 10 different classes.

Procedures for Use: In the study reported by Dopyera, raters observed the children for two 2-minute periods and recorded their observations immediately afterward. Their task was simply to indicate the occurrence or absence of specified events. Generally, 25 to 30 observations were recorded for each half-day session.

These records were date-time coded and transferred to tally sheets for data reduction. This involved converting the data to proportions to adjust for variations in the number of observations made. Proportion of item occurrence, or base rate, then, was the unit of analysis used.

Comments: Although reliabilities are somewhat low and two factors are uninterpretable, this observation procedure has several theoretical and practical advantages. Because it allows the researcher to record the child's actual encounters, this system is free from the assumption that the classroom environment is the same for all children. Information about the environmental factors confining or supporting each child's behavior is particularly valuable for purposes of program evaluation. Furthermore, the system can be used in almost any setting by minimally trained and relatively inexperienced observers.

Reference: Dopyera, J., *Assessing the micro-environments of individual preschool children.* Syracuse, New York: Head Start Evaluation Project, Syracuse University, 1969.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Miller-Hughes System (Miller-Hughes #23)

AUTHOR: George L. Miller, Marie Hughes
AUTHOR/PUBLISHER ADDRESS: G. L. Miller, Lesley College, Cambridge, Massachusetts

CRITERIA EVALUATED: 1, 3, 8, 14, 16, 17

SUBJECT OF OBSERVATION

____ TEACHER AND PUPILS
x TEACHER ONLY
____ PUPIL ONLY
____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
x 1 ONLY
____ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

____ LIVE
____ LIVE WITH CODING EQUIPMENT
x VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
x AFFECTIVE
x COGNITIVE
____ PROCEDURE OR ROUTINE
____ ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
____ TOPIC CHANGE
____ SPEAKER CHANGE
____ TIME UNIT
____ TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
x RESEARCH
x TRAINING
____ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 417-419.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Miniaturized Total Interaction Analysis System (Mini Total Interact #24)

AUTHOR: Herbert K. Heger
AUTHOR/PUBLISHER ADDRESS: University of Kentucky, Lexington, Kentucky

CRITERIA EVALUATED: 7, 8, 10, 13, 15, 16

SUBJECT OF OBSERVATION

[✓ TEACHER AND PUPILS]
TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

1 ONLY
[✓ MORE THAN 2 PEOPLE IN CLASSROOM]

COLLECTION METHODS

[✓ LIVE]
LIVE WITH CODING EQUIPMENT
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

[✓ AFFECTIVE]
COGNITIVE
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

[✓ CATEGORY CHANGE]
TOPIC CHANGE
SPEAKER CHANGE
[✓ TIME UNIT]
TIME SAMPLE
[✓ RATING SCALE]
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

[✓ RESEARCH]
TRAINING
EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 261-263.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Modified Category System (Modified Category #25)

AUTHOR: Edmund J. Amidon
AUTHOR/PUBLISHER ADDRESS:

CRITERIA EVALUATED: 8, 17, 19

SUBJECT OF OBSERVATION

[✓ TEACHER AND PUPILS]
TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

1 ONLY
[✓ MORE THAN 2 PEOPLE IN CLASSROOM]

COLLECTION METHODS

[✓ LIVE]
LIVE WITH CODING EQUIPMENT
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

[✓ AFFECTIVE]
[✓ COGNITIVE]
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

[✓ CATEGORY CHANGE]
TOPIC CHANGE
SPEAKER CHANGE
[✓ TIME UNIT]
TIME SAMPLE
RATING SCALE
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

[✓ RESEARCH]
[✓ TRAINING]
EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 121-123.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Nonverbal Interaction Analysis (Nonverbal Interaction #26)

AUTHOR: Peggy Amidon
AUTHOR/PUBLISHER ADDRESS: Paul S. Amidon and Associates, Minneapolis, Minnesota

CRITERIA EVALUATED: 9, 13, 16, 17, 19

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
____ TEACHER ONLY
____ PUPIL ONLY
____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
____ LIVE WITH CODING EQUIPMENT
x VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
____ PROCEDURE OR ROUTINE
x ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
____ TOPIC CHANGE

____ SPEAKER CHANGE
x TIME UNIT
____ TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
____ TRAINING
____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Page 464-466.

Name: NONVERBAL INTERACTION ANALYSIS

Author: Peggy Amidon

Type of Measure: Observation coding system

Availability: Amidon, P., *Nonverbal interaction analysis.* Minneapolis, Minnesota: Paul S. Amidon and Associates, 1971.

Description: Nonverbal Interaction Analysis (NVIA) provides a method of systematically observing and recording nonverbal behavior in the classroom. Designed to parallel the categories of verbal behavior in Flanders' Interaction Analysis (and capable of accomodating the 30 expanded IA categories), NVIA codes the following classroom dimensions, each of which affects teacher/pupil interaction.

I. Room Arrangement
II. Materials
III. Nonverbal Behaviors
IV. Activities

The first and second dimensions involve classroom setting—the physical arrangement of desks, chairs, bulletin boards, etc., and the presence or absence of certain materials, supplies, and equipment. The third dimension is concerned with the teacher's nonverbal behavior—gestures, facial expressions, posture, and so forth. The fourth dimension combines the previous 3 by focusing on the use of materials, by the teacher and the pupils, in conjunction with nonverbal behaviors.

Sample Items: Specific aspects of each NVIA dimension are coded pictorially. For example, under Dimension I (Room Arrangement) the chalkboard is symbolized by the following figure: ▭. Or, under Dimension II (Materials and Supplies), printed items, such as textbooks, workbooks, and newspapers, are represented by the symbol, 📖 . These symbols are further specified by the 10 numerical codes used in Interaction Analysis: (1) Accepts Feeling; (2) Praises or Encourages; (3) Accepts Student Ideas; (4) Asks Questions; (5) Lecturing; (6) Giving Directions; (7) Criticizing or Justifying Authority; (8) Student Talk—Response; (9) Student Talk—Initiation; (10) Silence or Confusion. By combining symbols from Dimensions I and II with Interaction Analysis category numbers, the *contents* of furniture or material may be described, as illustrated below.

$\boxed{4}$ symbolizes a chalkboard containing questions.

$\boxed{5}$ symbolizes a chalkboard containing lecture-type information.

📖6 symbolizes printed material containing directions or instructions.

symbolizes printed material written by the student in response to an assignment or request made by the teacher (e.g., an essay or written answers to questions).

If a more precise description of content is desired, the Dimension I and Dimension II symbols can be combined with an *expanded* Interaction Analysis category number and code letter:

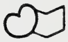
symbolizes printed material containing convergent-type questions.

Items on the NVIA Dimensions III and IV are also coded using combinations of symbols and Interaction Analysis categories. For example, on Dimension III (Nonverbal Behaviors), use of the hands to give directions is indicated by the following figure:

. Recording items on Dimension IV (Activities) is a bit more complex, since this aspect incorporates symbols from all of the preceding dimensions and distinguishes verbal from nonverbal use of material. Thus,

5 represents the verbal use of a book during a lecture,

while

 represents the use of a gesture to hold a book.

Reliability, Validity, and Norms: The NVIA manual (Amidon, 1971) reports no reliability, validity, or normative data.

Procedures for Use: Data collected on Dimensions I and II of the NVIA are recorded *prior* to observing the actual teacher/pupil interaction. The physical arrangement of the classroom and the materials present can be coded before class starts. Items on Dimensions III and IV, however, are sequentially recorded *during* classroom interaction. Immediate (live), delayed (videotaped), or a combination of immediate and delayed observation may be used. It is generally recommended that behaviors be recorded in 3-second units over a 20-minute observation period, though these specifications may be modified to accomodate unusual circumstances related to class size, students' ability, and classroom setting. Such circumstances, along with the purposes of observation, also determine the number of coders employed per classroom or per videotape. All or as few as 1 of the NVIA dimensions can be recorded—with or without the Interaction Analysis and expanded Interaction Analysis categories—depending on the user's needs and objectives.

Amidon (1971) explains NVIA recording procedures in great detail, illustrating the coding symbols for each dimension and their modification by Interaction Analysis and expanded Interaction Analysis categories. Practice exercises are also provided for prospective coders.

Comments: Because it incorporates Flanders' Interaction Analysis categories, Nonverbal Interaction Analysis is capable of providing a complete picture of classroom activity (both verbal and nonverbal). Yet it may also be used to assess only one or a few dimensions of teacher/student interaction. The NVIA is

adaptable to a number of situations and purposes and offers varying levels of specificity. The manual is very thorough in describing procedures for using the system. Although psychometric data are not reported, the NVIA, admittedly in an "infant stage," is based on an already researched and widely used observation instrument.

Reference: Amidon, P. *Nonverbal interaction analysis.* Minneapolis, Minnesota: Paul S. Amidon and Associates, 1971.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Observation Schedule and Record Form No. 4 (OSCAR 4V) (OSCAR 4V #27)

AUTHOR: Donald M. Medley
AUTHOR/PUBLISHER ADDRESS: Educational Testing Service, Princeton, New Jersey

CRITERIA EVALUATED: 8, 12, 16, 17, 19

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
____ TEACHER ONLY
____ PUPIL ONLY
____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
____ LIVE WITH CODING EQUIPMENT
____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
x PROCEDURE OR ROUTINE
____ ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
____ TOPIC CHANGE

x SPEAKER CHANGE
____ TIME UNIT
____ TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
____ TRAINING
____ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Page 156-157.

Name: OBSERVATION SCHEDULE AND RECORD FORM No. 4 (OScAR 4V)

Authors: Donald M. Medley, J. T. Impellitteri, and L. H. Smith

Type of Measure: Observation coding system

Availability: Dr. D. M. Medley, Chairman, Department of Research Methodology, Ruffner Hall, University of Virginia, Charlottesville, Virginia 22903.

Description: OScAR 4V is a system for classifying teacher verbal behaviors observed in the classroom. The focus of this recording system is on the teacher rather than the student. Content of teacher verbal behaviors is coded, while student verbalizations are merely tallied. OScAR 4V is primarily a research tool, consisting of 50 interrelated categories which are presented in flow-chart form. The flow chart guides the observer in classifying teacher statements. Tallies are placed on a separate recording form which contains all end points represented on the flow chart.

Sample Items: OScAR 4V codes classroom behavior under 3 major categories: (1) Teacher Statements; (2) Substantive Interchanges (teacher/student interaction focused on lesson content); and (3) Non-Substantive Interchanges (teacher/student interaction involving subjects other than lesson content). Each of these categories is further subdivided so that behaviors may be more specifically identified. Teacher Statements, for example, subsumes the following codes:

Teacher Statements

I. Affective	II. Procedural	III. Substantive
A. Considering (CSN)	A. Directing (DRC)	A. Informing (INF)
B. Rebuking (RBK)	B. Describing (DSC)	B. Problem Structuring (PRB)

Additional categories are used to indicate whether a verbalization is an Initial Statement (I) or a Continuing (C) remark.

Reliability: Reliability estimates are provided in a study described by Medley and Hill (undated), based on a sample of 72 intern teachers who were observed on OScAR 4V by 6 university supervisors. A reliability coefficient was calculated from an analysis of variance design for each OScAR 4V frequency count and contrast score. (A raw OScAR record is comprised of 42 frequency counts, while a scored record consists of 42 contrast scores, which are linear functions of these frequencies.) Estimated stability coefficients for contrast scores ranged from .00 to .85, with a median of .59. The range for frequencies was .00 to .84, with a median of .50. In general, the categories which proved least reliable were those with low frequencies, suggesting that instability may have been caused by the fact that certain behaviors were rarely observed.

Validity: No data indicating validity could be found. However, the flow chart which serves as a framework for OScAR 4V suggests logical category construction.

Norms: No normative data are available.

Procedures for Use: OScAR 4V requires no special equipment and can be easily learned. The observer uses the flow chart in categorizing classroom verbalizations. For example, in classifying a Teacher Statement, the observer refers to the flow chart, which asks if the statement called for a response. If the answer is no, the flow chart then asks if the statement had affective content. If the answer is no, the observer is guided to the next cell which asks if the statement had substantive content. If the answer is yes, the observer asks if it was a problem structuring statement. If so, he then determines whether it was an initiating or continuing remark.

Behaviors are tallied as they occur on the form provided. For research purposes, the observation period usually lasts between 5 and 10 minutes. However, there is no set time limit. Because only teacher verbalizations are coded for content, the observer's task is not difficult.

Comments: The OScAR 4V is a sophisticated instrument which is easy to use. It requires only 1 person to collect and code data. However, because validity information is unavailable, it is difficult to determine whether this system actually measures what it intends to measure. Nevertheless, OScAR 4V is a well organized and widely used instrument.

References: Medley, D. M., and R. A. Hill, *Measurement properties of Observation Schedule and Record, Form 4V*, undated. (mimeo)

Simon, A., and E. G. Boyer, *Mirrors for behavior. An anthology of observation instruments. Volume 3.* Philadelphia: Research for Better Schools, 1970.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Provo Code for the Analysis of Teaching (Provo Code #28)

AUTHOR: Marie Hughes
AUTHOR/PUBLISHER ADDRESS: University of Arizona, Tucson, Arizona

CRITERIA EVALUATED: 1, 3, 8, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☐ LIVE
☐ LIVE WITH CODING EQUIPMENT
☒ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☐ COGNITIVE
☒ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE
☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☐ 1 PERSON
☒ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☒ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 305-308.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Pupil-Teacher Rapport Scale (Pupil-Teacher Rapport #29)

AUTHOR: J. Wayne Wrightstone
AUTHOR/PUBLISHER ADDRESS: Source in Wrightstone (1951)

CRITERIA EVALUATED: 9, 12, 16, 17

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
___ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
x AFFECTIVE
___ COGNITIVE
___ PROCEDURE OR ROUTINE
x ACTIVITY
x SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

___ CATEGORY CHANGE
___ TOPIC CHANGE

___ SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
x RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
x RESEARCH
___ TRAINING
___ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Done Mills, Ontario, 1977, Pages 472-473.

Name: PUPIL-TEACHER RAPPORT SCALE

Author: J. Wayne Wrightstone

Type of Measure: Observation rating scale

Availability: Wrightstone, J. W., "Measuring the social climate of a classroom." *Journal of Educational Research*, 1951, **44**, 341–351.

Description: Wrightstone (1951) developed the Pupil-Teacher Rapport Scale in an attempt to measure the social climate of classrooms. Assuming that he could economically obtain an estimate of classroom atmosphere by observing and rating important components of pupil-teacher rapport, Wrightstone constructed a scale incorporating the following categories: Pupil-Teacher Interaction Pattern; Degree of Social Interaction; Quality of Social Interaction; Interest; Enjoyment; Role Structure; Emotion of Teacher; Teacher Orders or Suggestions; Physical Tension of Group; and Emotional Tone of Pupil Group.

The scale appears to be appropriate for use in either elementary or secondary classrooms.

Sample Items: An observer rates the classroom according to each of the categories listed above, using a 3-, 4-, or 5-point scale similar to the following:

Category	*Rating Scale*
Enjoyment	1. No pleasure—signs of displeasure or aversion.
	2. Little pleasure—minimal enjoyment.
	3. Some pleasure—a few signs of enjoyment.
	4. Moderate pleasure—generally restrained, occasionally active enjoyment.
	5. Extreme pleasure—enthusiasm.
Teacher Orders or Suggestions	1. Teacher demands compliance.
	2. Teacher expects compliance.
	3. Teacher generally suggests rather than coerces.
	4. Teacher allows options, commanding only in emergencies.

Reliability: Two estimates of interrater agreement are provided by Wrightstone (1951). Trained observers who had used the scale for 5 to 10 observations made independent ratings of the same classroom and achieved 90% agreement on item ratings. In addition, total scores given to 25 teachers by 2 independent observers correlated .92.

Validity: In order to test the validity of the scale, supervisors were asked to make detailed, anecdotal notes, over a period of months, describing the rapport between the teacher and the children in 25 classrooms. These supervisory reports covered the same general categories as those on the observation scale. Both quantitative and qualitative analyses were made, comparing the data provided

by supervisors to that obtained on the rating scale. These comparisons convinced Wrightstone (1951) that the scale had a "high degree of correspondence (p. 435)" with the anecdotal records.

Norms: No normative data are provided by Wrightstone (1951).

Procedures for Use: In using the scale, the observer visits the classroom on several occasions. The rating form provides space for 3 different ratings to be completed on 3 separate occasions. Identification data (name of teacher, school, class or grade, number of pupils, date, name of observer, length of observation, topic of class activity, and group structure) are recorded prior to observation. The rater then checks the items on the scale which best describe the social climate of the classroom.

Comments: The assumption underlying the Pupil-Teacher Rapport Scale—that the degree of compatibility between the teacher and his pupils is an indicator of social climate in the classroom—appears to be sound. Evidence of reliability and validity is positive, though limited, and the observational categories represented on the scale cover the major aspects of classroom interaction. This instrument does, however, require a high degree of inference on the rater's part and therefore requires extensive training of observers to insure that they are interpreting behavioral categories and observed behaviors similarly.

References: Wrightstone, J. W., "Measuring the social climate of a classroom." *Journal of Educational Research*, 1951, **44**, 341-351.

Wrightstone, J. W., J. Justman, and I. Robbins, *Evaluation in modern education.* New York: American Book Co., 1956.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Performance Assessment Record for Teachers (PART #30)

AUTHOR: Bob Burton Brown

AUTHOR/PUBLISHER ADDRESS: Institute for the Development of Human Resources, College of Education, University of Florida, Gainesville, Florida

CRITERIA EVALUATED: 13, 14(?), 16(?), 17(?), 18(?), 19(?)

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☐ LIVE
☐ LIVE WITH CODING EQUIPMENT
☒ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☒ TRAINING
☒ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Page 158-159.

Name: PERFORMANCE ASSESSMENT RECORD FOR TEACHERS

Author: Bob Burton Brown

Type of Measure: Observation coding system

Availability: Brown, B. B. *Performance assessment record for teachers.* Gainesville, Florida: Institute for Development of Human Resources, College of Education, University of Florida, 1970.

Description: The Performance Assessment Record for Teachers (PART) provides a framework for systematically observing, describing, and assessing the classroom performance of teachers and pupils. Though this instrument codes a number of pupil behaviors, it is designed primarily for teacher self-assessment. Audio or video tape recordings or live observation of classroom performance may serve as the basis of this assessment.

PART consists of 22 items which describe what does and does not happen in classrooms. The observer records the presence or absence of a given behavior during a brief (usually 6-minute) working period, without indicating the number of times it may have occurred. PART items are drawn from several other observation instruments. The first 4 are borrowed from the Teacher Practices Observation Record, which measures the extent to which classroom activity promotes inquiry or reflective thinking. Items 5 through 15 are adapted from the Florida Taxonomy of Cognitive Behavior, which assesses the level of cognitive operations overtly used by pupils in the classroom. Items 16 through 19, extracted from the Taxonomy of Affective Behavior, represent higher affective processes, particularly valuing. The remaining 3 items, drawn from the research of Soar (1970), deal with classroom control. Brown (1970) points out that these 22 items represent only a limited cross section of several alternative approaches to the analysis of teaching/learning situations in the classroom. PART items are not intended to represent universally "good" or "bad" behaviors. They should be used only as a means of gaining insight into classroom processes and effective teaching procedures.

Sample Items: The 22 PART items include teacher and pupil behaviors of the following type:

1. Teacher initiates a learning activity focusing on a question or problem which is of real concern to the student.
2. Teacher initiates a learning activity which addresses a question or problem arising from textbook or class discussion.
3. Student interprets, explains, tells why, compares, or points out causal relationships.
4. Student alters judgment in light of new information.

Brown (1970) lists specific statements or behaviors which indicate the occurrence of each item. For example, item 3, above, is coded when the student makes a statement similar to any of the following:

a. A campfire goes out when you cover it with dirt because fire needs oxygen in order to burn.

b. These 3 words sound alike, but they are spelled differently.

c. The plants on the window sill are dead because the window was left open last night and the temperature fell below freezing.

Reliability, Validity, and Norms: Because PART items were extracted from several other observation systems, no reliability, validity, or normative data are provided by Brown (1970).

Procedures for Use: Brown (1970) gives clear instructions for recording observations. Each 30-minute observation period is composed of 5 separate 7-minute working intervals, each of which corresponds to one of 5 columns on the coding sheet. During working period 1, the rater places a check in column 1 alongside the behavior observed. This procedure is repeated for each of the 6-minute working periods. When the entire observation is over, the number of checks recorded in columns 1 through 5 are totalled for each item. Therefore, the score for each item reflects frequency of occurrence.

Comments: The Performance Assessment Record for Teachers offers a method of evaluating teachers on an everyday basis. According to the author, this test is not intended for research purposes, but instead for use in the field as a self-assessment tool. Although PART covers limited content, such a restriction is necessary in an instrument designed for practical use by teachers. The absence of psychometric information, however, does weaken this measure in spite of the fact that its items are drawn from other, established observation systems.

References: Brown, B. B., *The experimental mind in education.* New York: Harper & Row, 1968.

Brown, B. B., *Instructional guide for the Performance Assessment Record for Teachers (PART).* Gainesville, Florida: Institute for Development of Human Resources, College of Education, University of Florida, 1970.

Soar, R. S., *Interim report, evaluation of Follow-Through model implementation.* Gainesville, Florida: Institute for Development of Human Resources, College of Education, University of Florida, 1970.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Revised Verbal Interaction Category System-Science (Rev. Verb. Interact. #31)

AUTHOR: Elizabeth Hunter
AUTHOR/PUBLISHER ADDRESS: Hunter College, New York, New York

CRITERIA EVALUATED: 8, 16, 17, 19

SUBJECT OF OBSERVATION

- ☒ TEACHER AND PUPILS
- ☐ TEACHER ONLY
- ☐ PUPIL ONLY
- ☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

- ☐ 1 ONLY
- ☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

- ☒ LIVE
- ☐ LIVE WITH CODING EQUIPMENT
- ☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

- ☒ AFFECTIVE
- ☐ COGNITIVE
- ☒ PROCEDURE OR ROUTINE
- ☒ ACTIVITY
- ☐ SOCIOLOGICAL STRUCTURE
- ☐ OTHER

SETTINGS

- ☒ CLASSROOM, ANY SUBJECT
- ☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

- ☒ CATEGORY CHANGE
- ☐ TOPIC CHANGE
- ☐ SPEAKER CHANGE
- ☐ TIME UNIT
- ☐ TIME SAMPLE
- ☐ RATING SCALE
- ☐ OTHER

OBSERVER(S) REQUIRED

- ☐ 1 PERSON
- ☒ TEAM OF 2 PERSONS
- ☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

- ☒ RESEARCH
- ☒ TRAINING
- ☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 309-312.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Social Weather Rating Scales (Social Weather #32)

AUTHOR: James E. Simpson
AUTHOR/PUBLISHER ADDRESS: Source in Simpson (1963)

CRITERIA EVALUATED: 12, 14, 16, 17

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☒ AFFECTIVE
☐ COGNITIVE
☐ PROCEDURE OR ROUTINE
☐ ACTIVITY
☒ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Page 480-481.

Name: THE SOCIAL WEATHER RATING SCALES

Author: James E. Simpson

Type of Measure: Observer rating scale

Availability: Simpson, J. E., "A method of measuring the social weather of children," in R. G. Barker (ed.), *The stream of behavior*. New York: Appleton-Century-Crofts, 1963.

Description: The Social Weather Rating Scales were designed to measure "social weather," that is, "the over-all social treatment given a particular individual in a behavior setting (Simpson, 1963, p. 219)." Based in part on the Fels Parent Behavior Rating Scales, this instrument assesses 9 specific factors which may be grouped to form 3 general dimensions of social weather. Each dimension and its component scales are listed below:

Dimension	*Scales*
1. Warmth (emotional tone of social weather)	1. Acceptance 2. Affectionateness 3. Approval
2. Tendance (direct efforts made for the child)	1. Attention 2. Assistance 3. Communication
3. Indulgence (the freedom allowed the child)	1. Adaptation 2. Privilege 3. Choice

The scales may be used to assess an individual child's treatment by the group, the teacher's behavior toward children, or, combining the ratings of individual children, to obtain an aggregate picture of group atmosphere.

Sample Items: Each bipolar scale is arranged vertically and punctuated by 5 objectives, which become progressively negative. The child's rating is indicated by checking the appropriate spot on the scale. Following each adjective is a brief definition, clarifying for the rater the conditions which warrant a check mark at that particular point on the scale. The Choice scale, for example, is similar to that below.

CHOICE: the freedom of choice or option allowed the child.

- *Free:*	Few or no attempts to make choices for the child.
- *Liberal:*	The child's choices are controlled occasionally, sporadically, or superficially.
- *Limited:*	The child is given some choice or offered alternatives that are indirectly coercive.
- *Controlled:*	Compliance is usually demanded of the child.
- *Restricted:*	The child has no choice. Absolute obedience is required.

Reliability: Interobserver and test/retest reliabilities are reported by Simpson (1963) for 2 observational situations. In Situation A, raters observed a child and his parents at dinner; in Situation B they observed the same child on the school playground at recess. Situation A was judged moderately difficult to rate, while Situation B was considered socially chaotic and thus very difficult to rate. The ratings of 4 observers correlated .83 for Situation A and .38 for Situation B, using Kendall's coefficient of concordance.

Test/retest estimates, over a 6-week interval, were computed on: (1) the investigator's ratings; (2) the investigator's ratings and median ratings of 3 untrained observers; and (3) the investigator's ratings and the average ratings of 3 untrained observers. Kendall's coefficients of concordance were, respectively: .98 for Situation A and .85 for Situation B; .87 (A) and .71 (B); and .91 (A) and .69 (B). Coefficients obtained for Situation A were significant at the .001 level, for Situation B at the .01 or .02 level.

Validity: Data describing the validity of the Social Weather Rating Scales are scarce. Findings reported by Simpson (1963), however, indicate that: the social weather of preschool children (measured on these scales) differs from that of elementary school children in shared behavior settings; social weather varies for each child from setting to setting; and variance for preschool children is smaller within the home and greater outside the home than that for school children.

Norms: Normative data are not reported (Simpson, 1963).

Procedures for Use: The Social Weather Rating Scales are used in conjunction with very explicit behavior setting records described by Simpson (1963). The rater does not observe the child directly but instead rates his social weather using an episodic account of his interactions, recorded previously. He rates the child's treatment by marking an X at that point on the scale which best describes the variable in question. Procedures are outlined for recording variations or exceptions in the behavior of significant individuals interacting with the child.

The rater must attempt to make "objective" judgments, relying on his own perception of the overall behavior toward the child—and not on the child's (or anyone else's) apparent perception of it. That is, if the teacher treats the child in a neutral way, this behavior should be rated "neutral," even though the child may feel that it is cold and rejecting.

Comments: The Social Weather Rating Scales offer a technique for assessing the atmosphere of a setting or the overall treatment received by an individual. Though validity and normative data are inadequate, these scales can be recommended for research use. They are particularly appropriate for research assessing the effect of environment on behavior.

Reference: Simpson, J. E., "A method of measuring the social weather of children," in R. G. Barker (ed.), *The stream of behavior*. New York: Appleton-Century-Crofts, 1963.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Student Activity Profile (Student Activity #33)

AUTHOR: Fred K. Honigman, James Stephens
AUTHOR/PUBLISHER ADDRESS: F. K. Honigman, Jefferson Medical College,
Philadelphia, Pennsylvania

CRITERIA EVALUATED: 9

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☐ TEACHER ONLY
☒ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM
☒ POINT-TIME SAMPLE*

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☐ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☐ CLASSROOM, ANY SUBJECT
☒ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 297-299.

* EACH STUDENT, BUT ONLY ONE AT A TIME

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Student Behavior Observation System (Student Behaviour #34)

AUTHOR: Hugh V. Perkins
AUTHOR/PUBLISHER ADDRESS: University of Maryland, College Park, Maryland

CRITERIA EVALUATED: 9

SUBJECT OF OBSERVATION

- ☐ TEACHER AND PUPILS
- ☐ TEACHER ONLY
- ☒ PUPIL ONLY
- ☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

- ☐ 1 ONLY
- ☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

- ☐ LIVE
- ☒ LIVE WITH CODING EQUIPMENT
- ☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

- ☒ AFFECTIVE
- ☐ COGNITIVE
- ☐ PROCEDURE OR ROUTINE
- ☒ ACTIVITY
- ☐ SOCIOLOGICAL STRUCTURE
- ☐ OTHER

SETTINGS

- ☒ CLASSROOM, ANY SUBJECT
- ☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

- ☒ CATEGORY CHANGE
- ☐ TOPIC CHANGE
- ☐ SPEAKER CHANGE
- ☐ TIME UNIT
- ☒ TIME SAMPLE
- ☐ RATING SCALE
- ☐ OTHER

OBSERVER(S) REQUIRED

- ☒ 1 PERSON
- ☐ TEAM OF 2 PERSONS
- ☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

- ☒ RESEARCH
- ☐ TRAINING
- ☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 477-479.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Student Observational Form (Student Obs. Form #35)

AUTHOR: C. M. Lindvall
AUTHOR/PUBLISHER ADDRESS: Learning Research and Development Center, University of Pittsburgh, Pittsburgh, Pennsylvania

CRITERIA EVALUATED: 9, 18

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☐ TEACHER ONLY
☒ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☐ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM
☒ POINT-TIME SAMPLE

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☐ AFFECTIVE
☐ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☒ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☐ CLASSROOM, ANY SUBJECT
☒ CLASSROOM, SPECIFIC SUBJECT (I.E.P)

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE
☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 353-356.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Sequential Analysis of Verbal Interaction (SAVI #36)

AUTHOR: Yvonne Agazarian and Anita Simon
AUTHOR/PUBLISHER ADDRESS: Savi Associates, Philadelphia, Pennsylvania

CRITERIA EVALUATED: 7, 14, 16, 17

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
____ TEACHER ONLY
____ PUPIL ONLY
x SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

____ LIVE
____ LIVE WITH CODING EQUIPMENT
x VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
____ COGNITIVE
____ PROCEDURE OR ROUTINE
x ACTIVITY
____ SOCIOLOGICAL STRUCTURE
____ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
____ TOPIC CHANGE

____ SPEAKER CHANGE
x TIME UNIT
____ TIME SAMPLE
____ RATING SCALE
____ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
x TRAINING
x EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Page 474-476.

Name: SEQUENTIAL ANALYSIS OF VERBAL INTERACTION

Authors: Yvonne Agazarian and Anita Simon

Type of Measure: Observation coding system

Availability: Agazarian, Y., and A. Simon, *Sequential analysis of verbal interaction.* Philadelphia, Pennsylvania: SAVI Associates, 1974.

Description: Sequential Analysis of Verbal Interaction (SAVI), originally developed in 1960 and revised in 1974, is a system for analyzing verbal communication among people. SAVI is based on the assumption that the way in which something is said influences what is heard. This instrument provides a "mental snapshot" of what happens when people talk and predicts what is likely to happen next.

The SAVI categorizes verbal behavior as either approaching or avoiding problems to be solved. Approach Behaviors facilitate the transfer of information, while Avoidance Behaviors tend to distort or confuse information. In addition to these 2 types of behavior, the SAVI also describes Contingent Behaviors, the direction and effect of which are not always obvious. The impact of Contingent Behaviors on problem solving is, as the name implies, contingent upon the interactional context.

The SAVI also classifies messages as either Person, Factual, or Orienting. Person Messages address the person rather than what is said. Factual Messages contain the kinds of information that are basic to data input and data processing, and Orienting Messages determine the level of abstraction and sequence of context characterizing communication. Each of these types of messages is associated with 1 of the 3 categories of behavior, producing 9 classes of communication: (1) Person Avoidance Messages; (2) Person Approach Messages; (3) Person Contingent Messages; (4) Factual Avoidance Messages; (5) Factual Approach Messages; (6) Factual Contingent Messages; (7) Avoidance Orienting Messages; (8) Approach Orienting Messages; and (9) Contingent Orienting Messages. When communication is categorized by this procedure, patterns of verbal behavior emerge. For example, if the frequency of Avoidance Messages is high, the overall pattern indicates defensive communication. In such a situation it is unlikely that information relevant to some objective goal or target will be transferred in ways that lead to effective problem solving.

The authors believe that identifying patterns of communication is the first step toward influencing them. They designed SAVI to allow individuals to gain practical and theoretical knowledge of *how* to say *what* in order to resolve the problems inherent in communicating. Such an instrument has obvious implications for teacher training and analysis of classroom interaction.

Sample Items: The 9 classes of communication are characterized by behaviors of the following type.

Person Avoidance Message: self-defensive behavior
Person Contingent Message: personal history
Person Approach Message: personal support
Factual Avoidance Message: small talk
Factual Contingent Message: factual question
Factual Approach Message: factual answer
Orienting Avoidance Message: being obscure
Orienting Contingent Message: opinion
Orienting Approach Message: building on ideas

Reliability: Reliability estimates are reported by Agazarian (1968), who observed a group of people and taped their voices, later categorizing the recorded verbal behaviors on the SAVI. Intercoder agreement, calculated for 10-minute periods, ranged from 87 to 96%. During a later phase of this study, reliability was checked by obtaining judgments from two separate raters on a systematic sampling of SAVI classifications. Agreement between judges and agreement with the initial coding was 100%.

Validity: Agazarian (1968) selected 2 external criteria against which to validate the SAVI: targetness (the goal orientation of the group) and convergence (continuity of topic). Communication was viewed as convergent when the topic of a message was the same as that of the preceding message, and as divergent when a new topic was introduced. Subjects observed were a group of people attending a conference to solve a specific problem. Their communication was categorized on the SAVI. Incidents of verbal behavior were also independently judged according to the selected criteria, i.e., as either on or off target, and convergent or divergent. Fisher's discriminate function yielded data which confirmed the ability of the SAVI categories to discriminate targetness and convergence at less than the .01 level of probability. A *chi* square statistic showed that 9 specific categories were significantly related to targetness, 3 in an off-target direction and 6 in an on-target direction. Three categories related significantly to convergence and seven to divergence. These findings, along with the overall distribution of the categories, revealed a pattern of cross-purpose talk. Since the group observed had a cross-purpose goal, it was expected that both interpersonal maintenance and interpersonal conflict behaviors would be avoided, as they were.

Norms: No normative data are reported (Agazarian, 1968; Agazarian and Simon, 1974).

Procedures for Use: In using the SAVI, verbal behavior is recorded on tape (usually in 10-minute samples) and later transcribed. Listening to the tape, a coder classifies the verbal interaction by labeling the transcribed statements according to the appropriate SAVI category. The coded transcript is scored in terms of frequency of occurrence of various kinds of behavior.

Comments: The SAVI system provides a unique perception of communication using 9 classes of behavior which are well defined and illustrated. These 9 classes can accommodate all verbal behavior to provide an objective basis on which to analyze the effectiveness of communication. Additional research data, including further evidence of reliability and validity, should be obtained and reported. Since this instrument appears to have value for teacher training or classroom interaction assessments, data collected in educational settings would be of interest.

References: Agazarian, Y., *A theory of verbal behavior and information transfer.* Unpublished doctoral dissertation, Temple University, 1968.

Agazarian, Y., and A. Simon, *Sequential analysis of verbal interaction.* Philadelphia, Pennsylvania: SAVI Associates, 1974.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Social Interaction and Creativity in Communication (SICC #37)

AUTHOR: David L. Johnson
AUTHOR/PUBLISHER ADDRESS: Institute of Human Resources, Streator, Illinois

CRITERIA EVALUATED: 14, 16, 17, 19

SUBJECT OF OBSERVATION

_x_ TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

SETTINGS

_x_ CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

_x_ CATEGORY CHANGE
___ TOPIC CHANGE
___ SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
___ RATING SCALE
___ OTHER

NUMBER OF SUBJECTS OBSERVED

___ 1 ONLY
_x_ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

_x_ LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

OBSERVER(S) REQUIRED

_x_ 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

CATEGORY DIMENSIONS

_x_ AFFECTIVE
_x_ COGNITIVE
___ PROCEDURE OR ROUTINE
_x_ ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

USES REPORTED BY AUTHOR

_x_ RESEARCH
___ TRAINING
___ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 477-479.

Name: SOCIAL INTERACTION AND CREATIVITY IN COMMUNICATION SYSTEM

Author: David L. Johnson

Type of Measure: Observation coding system

Availability: Johnson, D. L., *Social Interaction and Creativity in Communication System. Coding manual.* Streator, Illinois: Institute of Human Resources, 1975.

Description: The Social Interaction and Creativity in Communication System (SICCS) was designed to assess the social and environmental conditions facilitating, hindering, or otherwise affecting creative performance. The product of extensive and continuing developmental work, SICCS is based on a well defined, social/psychological conceptualization of creativity. This conceptual model focuses on verbal acts in the context of social interaction. According to Johnson (1975), "verbal acts may be creative, and creative *processes* may be examined sequentially in relation to antecedent and subsequent speaker acts (p. 6)." Creativity is viewed as an unexpected, original verbal act emitted spontaneously by a speaker in social interaction and affected by the verbal acts of antecedent speakers. The verbal act (a phrase, clause or simple sentence) and the interact (a verbal act followed by at least 1 verbal act emitted by another speaker) serve as the units of analysis employed by the SICCS. They are observed and coded according to the following categories:

I. Antecedent Modes of Verbal Activity
  A. Appraisal
    1. Positive Judgment
    2. Negative Judgment
  B. Prescriptive
    1. Direction
    2. Standard
  C. Informational
    1. Certain
    2. Uncertain
  D. Questioning
    1. Narrow
    2. Broad

II. Subsequent Student Performance
  A. Self-Reference
    1. Positive
    2. Negative
  B. Self-Initiated
    1. Volunteering
    2. Redirecting
    3. Interrupting

C. Productivity
  1. Productive
  2. Elaboration
  3. Construction
  4. Reinterpretation
D. Quantity
  1. Length
  2. Number

Sample Items: Johnson (1975) defines the categories listed above and provides numerous examples of verbal acts typically recorded under each one. For instance, a self-referenced verbal act is " one in which the speaker makes explicit reference to his own characteristics (p. 21)." It is coded positively if the speaker asserts responsibility for the content (e.g., "I think people usually do what is expected of them."). However, when the speaker repudiates or denies responsibility for the content of a verbal act (e.g., "I didn't say that."), it is coded negatively.

Reliability: Using a Kappa (K) coefficient of agreement, Johnson (1971) calculated mean interrater reliabilities for all SICCS categories based on sixth- and eleventh-grade lesson transcripts. These ranged from .85 to .93 for the antecedent mode categories and from .87 to 1.00 for the student performance categories. Mean percentages of intercoder agreement computed on 10 bilingual transcripts from kindergarten and first- through third-grade classes ranged from 95.6 to 97.6% for the mode categories, and from 91.5 to 100.0% for the performance categories (Johnson, 1974).

Validity: Although the SICCS is a relatively new instrument, various forms have been employed in a number of studies, and evidence of validity is accumulating. For example, Smith (1972), using the SICCS to observe 6 small group sessions in a drug-abuse treatment center, found that uncertain rather than certain informational acts and broad rather than narrow questions were positively related to a greater number of subsequent verbal acts. Childers (1974), studying open and closed school systems, observed fourth- through sixth-grade classrooms and reported a higher incidence of self-initiated, positive self-referenced, and longer interactions in the open rather than the closed environments.

Studies by Johnson have also produced results in the expected directions. Observing 4 classrooms in a small, coeducational community college, he found the teachers' use of broad questioning, more than any other mode category, to be significantly and consistently associated with student creativity recorded on the SICCS performance categories. Also, the teachers' uncertain informational acts were more consistently related to student talents than were the prescriptive or appraisal modes (Johnson, 1972). Similarly, in an analysis of 24 taped social studies lessons from sixth- and eleventh-grade classrooms, Johnson (1973) found that both broad questioning and uncertainty in teacher informational

acts were associated with a higher quantity and quality of student verbal performance. Teacher appraisals, on the other hand, appeared to terminate or inhibit student participation.

Norms: No comprehensive normative data are provided in the SICCS *Coding Manual* (Johnson, 1975). However, classroom interaction outcomes for various groups are reported (as frequencies and percentages) in his individual studies (Johnson, 1972; 1973; 1974).

Procedures for Use: Student/teacher interaction can be studied on the SICCS using: (1) direct observation; (2) direct observation and audio tape recording; or (3) direct observation and audio/video tape recording. When either of the last 2 procedures is employed, the tape, rather than the live interaction, is coded. Johnson (1975) provides detailed instructions, including an explanation of codes, the sequence of coding steps, and examples of tape transcripts and completed coding sheets, for all 3 procedural alternatives.

Comments: The Social Interaction and Creativity in Communication System is based on the idea that student creativity is affected by the social/psychological conditions present in the classroom. Though carefully developed and revised according to an explicit theoretical model, this instrument is more complex than many of the observation coding systems available, and its manual is less clear than it might be. However, the SICCS demonstrates good reliability and validity and, when used by trained observers, is a useful research tool.

References: Childers, V., *Open-closed school system expectations and observed pupil behavior*. Unpublished master of arts thesis, Baylor University, 1974.

Johnson, D. L., *Classroom interaction and student creativity*. Unpublished doctoral dissertation. University of Missouri, 1971.

Johnson, D. L., *Teacher and student classroom interaction and student verbal creativity*. Paper presented at the annual meeting of the Southwestern Social Science Association, San Antonio, Texas, March 1972.

Johnson, D. L., "A conceptual model of teacher and student classroom interaction and observed student verbal creativity." *Psychology in the Schools*, 1973, **10**, 475–481.

Johnson, D. L., *Teacher-pupil interaction in bilingual elementary school classrooms*. Paper presented at the annual meeting of the Southwestern Social Science Association, Dallas, Texas, March 1974.

Johnson, D. L., *Social Interaction and Creativity in Communication System. Coding manual*. Streator, Illinois: Institute of Human Resources, 1975.

Smith, R., *The effects of various verbal modalities on the quantity of subsequent verbal acts*. Senior research paper, Missouri Southern State College, 1972.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Spaulding Teacher Activity Rating Schedule (STARS) (Forms Battery With CASES - Spaulding) (STARS #38)

AUTHOR: Robert L. Spaulding

AUTHOR/PUBLISHER ADDRESS:
Institute for Research in Child Development, Room
200, School of Education, San Jose University, 125
South 7th Street, San Jose, California 95192

CRITERIA EVALUATED: 1, 7, 12, 16, 17

SUBJECT OF OBSERVATION

___ TEACHER AND PUPILS
x TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

x 1 ONLY
___ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
x PROCEDURE OR ROUTINE
x ACTIVITY
x SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

___ CATEGORY CHANGE
___ TOPIC CHANGE

___ SPEAKER CHANGE
___ TIME UNIT
x TIME SAMPLE
___ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
x TRAINING
___ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 167-169.

Name: SPAULDING TEACHER ACTIVITY RATING SCHEDULE (STARS)

Author: Robert L. Spaulding

Type of Measure: Observation coding system

Availability: This instrument may be ordered from: Robert L. Spaulding, Institute for Research in Child Development, Room 200, School of Education, San Jose State University, 125 South 7th Street, San Jose, California 95192.

Description: The Spaulding Teacher Activity Rating Schedule (STARS) is an observation coding system designed to assess the cognitive instructional methods and affective and control strategies of teachers. The current form is a 1975 revision of an earlier edition reviewed in *Mirrors for Behavior* (Simon and Boyer, 1970). Both forms have been used in teacher training to provide feedback to prospective teachers about their cognitive and affective styles. The revised form has also been used, along with its companion instrument, Coping Analysis Schedule for Educational Settings (CASES), in a study relating teacher behaviors and student outcomes (Spaulding, 1975b).

Sample Items: STARS consists of several molar categories, under which are subsumed a number of subcategories. These categories, listed below, are defined by Spaulding (1975b):

1. General
    a. Nontransactional behavior
    b. Disapproval with aversive stimuli
    c. Disapproval without aversive stimuli
    d. Withholding reinforcers when a student bids for attention
    e. Approval with positive affect present
2. Social and/or Motor Structuring and Restructuring
    a. Social and/or motor structuring
    b. Social and/or motor restructuring
3. Cognitive Structuring and Restructuring—unrelated to academic task
4. Concept Formation, Attainment, or Development by Discovery Methods
    a. Simple
    b. Complex
5. Concept Formation, Attainment, or Development by Deductive Methods
    a. Simple
    b. Complex
6. Concept Formation, Attainment, or Development by Transductive Methods
    a. Simple
    b. Complex

7. Concept Formation, Attainment, or Development by Expository Methods
    a. Telling—Simple
    b. Telling—Complex
    c. Rote Process—Simple
    d. Rote Process—Complex
8. Cognitive Structuring or Restructuring by Focusing Attention on Data
9. Concept Checking, Testing, Strengthening
    a. Asking for or eliciting recall—Simple
    b. Asking for or eliciting recall—Complex
    c. Asking for use or application—Simple
    d. Asking for use or application—Complex
10. Valuing
    a. Expressing
    b. Eliciting
11. Listening and/or Observing

Reliability: Preliminary work by Spaulding (1975a) indicates that split-half reliability is highly variable for some categories and more consistent for others. Routine recordings of interobserver agreement fall in the .80s and .90s.

Validity: STARS is based on a factor analytic study of 113 categories of teacher/student interaction observed in 21 elementary school classrooms. Three types of teacher variables recorded on STARS were significantly related to student performance and self-concept: (1) supportive teacher behaviors which served as rewards; (2) aversive behaviors which served as punishment; and (3) goal setting behaviors which served to clarify or structure the environment for the student. Other researchers have also found these variables to be significantly related to change in student behaviors (Sibley, Abbott, and Cooper, 1967, as cited by Simon and Boyer, 1970).

Norms: Norms can be derived by summing frequencies over several observation periods. An example of norm setting is provided by Simon and Boyer (1970).

Procedures for Use: STARS is generally used in conjunction with CASES. Together, the measures yield a picture of student/teacher interaction. Data are gathered by time sampling techniques, using 10-second intervals.

STARS yields a total score derived from a weighted interaction matrix using CASES and STARS. Frequencies from each cell of the matrix are weighted, added, and divided by the total number of tallies to produce an Overall STARS Coefficient.

Comments: STARS appears to be a relatively reliable and valid measure of the teacher's cognitive, affective, and control behaviors. STARS is a versatile system which can be used in conjunction with a variety of instruments, in a variety of settings, for both research and teacher training.

References: Simon, A., and E. G. Boyer, *Mirrors for behavior. An anthology of observation instruments.* Philadelphia: Research for Better Schools, 1970.

Spaulding, R. L., Personal communication, November 26, 1975. (a)

Spaulding R. L., *The Spaulding Teacher Activity Rating Schedule* (*STARS*). San Jose, California: Institute for Research in Child Development, School of Education, San Jose State University, 1975. (mimeo) (b)

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Teacher Behavior Observation System (Teacher Behaviour #39)

AUTHOR: Hugh V. Perkins
AUTHOR/PUBLISHER ADDRESS: University of Maryland, College Park, Maryland

CRITERIA EVALUATED: 4, 16, 17, 19

SUBJECT OF OBSERVATION

TEACHER AND PUPILS
[✓ TEACHER ONLY]
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
[✓ 1 ONLY]
MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

LIVE
[✓ LIVE WITH CODING EQUIPMENT]
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
[✓ AFFECTIVE]
COGNITIVE
PROCEDURE OR ROUTINE
ACTIVITY
SOCIOLOGICAL STRUCTURE
OTHER

SETTINGS

[✓ CLASSROOM, ANY SUBJECT]
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

[✓ CATEGORY CHANGE]
TOPIC CHANGE

SPEAKER CHANGE
TIME UNIT
[✓ TIME SAMPLE]
RATING SCALE
OTHER

OBSERVER(S) REQUIRED

[✓ 1 PERSON]
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
[✓ RESEARCH]
TRAINING
EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 473-475.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Instrument for Coding Teacher Verbal and Non-Verbal Behavior (Teacher Verb-NonVerb #40)

AUTHOR: E. Wayne Roberson
AUTHOR/PUBLISHER ADDRESS: Epic Evaluation Center, Tucson, Arizona

CRITERIA EVALUATED: 4, 11, 12, 16, 17, 19

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☐ LIVE
☐ LIVE WITH CODING EQUIPMENT
☒ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☒ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☒ RESEARCH
☒ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 513-516.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Teacher-Child Dyadic Interaction (Teacher-Child Dyadic #41)

AUTHOR: Jere E. Brophy, Thomas L. Good
AUTHOR/PUBLISHER ADDRESS: University of Texas, Austin, Texas

CRITERIA EVALUATED: 3, 8, 17, 19

SUBJECT OF OBSERVATION

x TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

___ 1 ONLY
x MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
x PROCEDURE OR ROUTINE
___ ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

x CATEGORY CHANGE
x TOPIC CHANGE

x SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
___ RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
___ TRAINING
___ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 191-197.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Waimon System (Waimon #42)

AUTHOR: Morton D. Waimon
AUTHOR/PUBLISHER ADDRESS: Illinois State University, Normal, Illinois

CRITERIA EVALUATED: 8,11, 16, 17, 19,

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☒ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☒ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☒ TRAINING
☐ EVALUATION

REFERENCE:

Mirrors for Behavior III: An Anthology of Instruments, Anita Simon and E. Gil Boyer, Communications Materials Center, Wyncote, Pennsylvania, 1974, Pages 639-641

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Wallen's Modification of Flanders' Interaction Analysis (Wallen's Flanders Mod. #43)

AUTHOR: Norman E. Wallen
AUTHOR/PUBLISHER ADDRESS: Source in Wallen (1966)

CRITERIA EVALUATED: 4, 7, 11, 17, 19

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☐ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☒ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☐ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☒ RESEARCH
☐ TRAINING
☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 173-174.

Name: WALLEN'S MODIFICATION OF FLANDERS' INTERACTION ANALYSIS

Author: Norman E. Wallen

Type of Measure: Observation coding system

Availability: Wallen, N. E. *Relationships between teacher characteristics and student behavior—Part III*. Salt Lake City, Utah: Department of Educational Psychology, University of Utah, 1966. (ERIC Document Reproduction Service No. ED 010 390)

Description: This observation procedure was developed in a study of teacher behavior and pupil change (Wallen, 1966). In designing this system, the author modified Flanders' Interaction Analysis (1962) in regard to focus, recording procedures, and number and nature of behavioral categories observed. The focus, for example, was limited to teacher behavior only. Also, the continuous, sequential recording which characterizes the Flanders system was abandoned in favor of a simple frequency approach. Finally, some of Flanders' categories which were inapplicable to or inappropriate for the first- and third-grade samples used by Wallen were altered, producing the behavioral classifications listed below.

1. Acknowledges student's raised hand
2. Praise and encouragement
2a. Nonverbal affiliation
3. Minimal reinforcement
4. Asking questions with the intent that the student answer.
5. Explaining or problem structuring
6. Academic control
7. Personal control
7m. Moralizing by teacher
8. Hostility and reprimands
8x. Ignores child's behavior when child is attempting to get teacher's attention.

In using the system, the observer may record a given teacher behavior in more than one category. For example, if the teacher praises a pupil in an attempt to control the others, the action would be tallied as both a 7 (personal control) and then as either a 2 or a 3, depending on the strength of the reinforcement given.

Sample Items: While many of Wallen's categories are self-explanatory, he provides examples of specific behaviors which are coded under each area. "Nonverbal affiliation," for example, includes physical contact with the student, such as putting an arm around him or her. "Minimal reinforcement" refers to comments such as, "uh huh," "okay," "right," etc.

Reliability: The records of 4 raters who had observed the same teachers at the same time, shortly before and midway through the major data collection period, where compared by Wallen (1966). Descriptive profiles for 2 teachers showed adequate agreement among observers. Correlations between observations recorded simultaneously by the 2 members of each recording team also showed good correspondence. These ranged from .85 to .96 for Team I ($N = 33$ teachers) and from .65 to .94 for Team II ($N = 34$ teachers).

The 2 observation teams were also compared with regard to overall mean frequencies in each category. They showed close agreement in all areas except problem structuring (Wallen, 1968).

Validity: Validity may be inferred from correlations between teacher behavior observed on this system and (1) other assessments of teacher classroom behavior and (2) measures of pupil behavior.

After all observations had been completed, Wallen (1966) instructed observers to Q-sort the sample of teachers according to 10 categories analogous to those on the observation system. Results of these Q-sorts were correlated with observational data, revealing a satisfactory and expected relationship between the two. For example, "praise, and encouragement" on the observation scale correlated positively with the Q-sort dimensions, "warmth" and "supportive behavior." Coefficients were .67 and .60, respectively, for the first grade, and .47 and .42 for the third grade. Similarly, "praise and encouragement" correlated negatively with "punitive behavior" on the Q-sort.

Relationships between teacher behavior and pupil behavior also generally supported the validity of the observation system. Measures of pupil behavior included the following: California Achievement Tests; Circles Test from the Torrance Tests of Creative Thinking; the Barron-Welsh Art Scales; a sociometric device; and a Pupil Questionnaire composed of items from the Sarason Test Anxiety Scale and Medley and Kline's Liking for School Test (Wallen, 1966).

Norms: Wallen (1966) presents observational data collected on 36 first-grade and 40 third-grade teachers in a large metropolitan school district in Salt Lake City.

Procedures for Use: In using this system, the observer sits in the classroom for five 30-minute periods and records teacher behaviors in the appropriate categories. No attempt is made to indicate the sequence of these behaviors. Both frequency and proportion (percentage) scores are derived from the data obtained.

Comments: Wallen's observation system offers an uncomplicated, easy-to-use procedure for assessing the affective behavior of teachers. It requires low inference and relatively limited training on the coders' part and demonstrates adequate agreement among observers.

Reference: Wallen, N. E., *Relationships between teacher characteristics and student behavior—Part III*. Salt Lake City, Utah: Department of Educational Psychology, University of Utah, 1966. (ERIC Document Reproduction Service No. ED 010 390)

Section B:

Instrument Summaries for Pupil Use

Table 2

Instruments and Criteria for Pupil Use

	Criteria																		
Instrument	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Class Activities Questionnaire										X			X	X			X		X
2. Diagnostic Teacher Rating Scale				X			X			X	X						X	X	
3. Illinois Teacher Evaluation Questionnaire																			
4. LIE						X					X	X		X	X				
5. Pupil Observation Survey				X			X			X									
6. Purdue Teacher Evaluation							X					X					X		
7. Students' Evaluation of Instruction	X					X				X	X	X			X				
8. Student Evaluation of Teaching I							X										X		
9. Student Evaluation of Teacher II										X							X		
10. Student Perception of Teaching Style																X			
11. Teacher Image Questionnaire																			

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Class Activities Questionnaire (#1)

AUTHOR: Joe M. Steele
AUTHOR/PUBLISHER ADDRESS: Dr. Robert Rosemier, 520 College View Court,
Northern Illinois University, DeKalb, Illinois, 60115

CRITERIA EVALUATED: 10, 13, 14, 17, 19

SUBJECT OF OBSERVATION

__x__ TEACHER AND PUPILS
_____ TEACHER ONLY
_____ PUPIL ONLY
_____ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

_____ 1 ONLY
__x__ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

__x__ LIVE (RETROSPECTIVE RATINGS)
_____ LIVE WITH CODING EQUIPMENT
_____ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

__x__ AFFECTIVE
__x__ COGNITIVE
__x__ PROCEDURE OR ROUTINE
_____ ACTIVITY
__x__ SOCIOLOGICAL STRUCTURE
_____ OTHER

SETTINGS

__x__ CLASSROOM, ANY SUBJECT
_____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

_____ CATEGORY CHANGE
_____ TOPIC CHANGE
_____ SPEAKER CHANGE
_____ TIME UNIT
_____ TIME SAMPLE
__x__ RATING SCALE
_____ OTHER

OBSERVER(S) REQUIRED

__x__ 1 PERSON (TEACHERS AND STUDENTS)
_____ TEAM OF 2 PERSONS
_____ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

_____ RESEARCH
_____ TRAINING
__x__ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 402-405

Name: CLASS ACTIVITIES QUESTIONNAIRE

Author: Joe M. Steele

Type of Measure: Rating scale

Availability: This instrument can be ordered from Dr. Robert Rosemier, 520 College View Court, Northern Illinois University, DeKalb, Illinois 60115.

Description: The Class Activities Questionnaire (CAQ) is a measure of instructional climate administered to both the student (in grades 7 and above) and the teacher. CAQ results obtained from students suggest *actual* classroom emphases, while those from the teacher indicate *intended* emphases. These 2 sets of data can be compared for purposes of research or teacher self-assessment.

Developed in 1969 as part of a large-scale evaluation of gifted classes throughout the state of Illinois, the CAQ was designed to incorporate 2 theoretical domains of instructional climate: Cognitive and Affective. Results of field testing and factor analysis indicated that these domains are represented on the questionnaire by the 5 dimensions and 20 factors listed below:

Dimensions	*Factors*
1. Lower Thought Processes	1. Memory
	2. Translation
	3. Interpretation
2. Higher Thought Processes	4. Application
	5. Analysis
	6. Synthesis
	7. Evaluation
3. Classroom Focus	8. Discussion
	9. Test/Grade Stress
	10. Lecture
4. Classroom Climate	11. Enthusiasm
	12. Independence
	13. Divergence
	14. Humor
	15. Ideas Valued
	16. Ideas Enjoyed
	17. Teacher Talk
	18. Homework
5. Student Opinions	19. Qualities
	20. Deficiencies

These factors are in turn represented by items describing (1) cognitive demands made of students and (2) affective classroom conditions.

Sample Items: The CAQ contains 25 items of the following type to which the

subject responds on a 4-point scale (from "Strongly Agree" to "Strongly Disagree"):

1. The student's main task is to remember or recognize factual material.
2. Students in this class are often required to decide whether things are good or bad, right or wrong, and to explain why.
3. Students are enthusiastic about and interested in class activities.
4. In this class there is very little laughing or kidding around.

Five additional CAQ items (3 of which are open-ended) request information about teacher activity, student preparation for class, and student preferences and opinions. The questionnaire also provides space for 20 optional, locally generated items.

Reliability: Though reliability figures are not reported for the most recent revision of the CAQ, estimates based on a preliminary version of the questionnaire are available (Steel, House, and Kerins, 1971). Using the Horst formula (1949), reliabilities were computed from within-class and between-class variances ($N = 131$ classes). These ranged from .76 to .88 for the CAQ dimensions and from .58 to .94 for the factors. Only 1 of 20 coefficients fell below .65.

Test/retest reliabilities, based on factor scores from 6 classes, were also computed for an early form of the CAQ. These figures, which ranged from .59 to .91, were viewed as tentative since the classes tested, being small and employing independent study modes, were considered atypical (Steele, House, and Kerins, 1971).

Validity: Studies conducted with preliminary versions of the CAQ offer some evidence of construct and concurrent validity. Factor analyses of CAQ results obtained from gifted classes substantially supported the logical construction of the instrument, though 2 factors, Divergence and Memory, proved to be weak. Generally, statistical components were closely related to the theoretical structure of the CAQ, with 8 of 10 logically paired items intact after empirical analysis (Steele, House, and Kerins, 1971). Using a normal rather than a gifted population, Wahlstrom (1971) analyzed the CAQ and confirmed its affective and cognitive dimensions, thus supplying further factorial validation of the scale.

Concurrent validity is suggested by differences (significant beyond the .001 level) between the CAQ subscores of average and gifted samples. Average classes as a group are seen as emphasizing 3 of the 7 cognitive processes represented on the scale, while gifted classes are seen as stressing 6. Only 8% of the average classes emphasized 4 or more cognitive levels, but 45% of the gifted classes emphasized from 4 to 6 levels.

On the affective dimensions of the CAQ, the contrast between average and gifted classes was also marked. Average classes were characterized by lack of enthusiasm, stress on grades and tests, and a greater proportion of time (75% to

90%) devoted to teacher talk. Gifted classes, on the other hand, stressed discussion and active student involvement. In almost one-third of these classes, teachers were seen as talking less than 40% of the time. Furthermore, gifted students reported less stress on tests and grades and a great deal more enthusiasm (Steele, House, and Kerins, 1971).

Norms: Extensive normative data are not reported (Steel, 1969; Steele, 1973). However, some comparative statistics (percentage of classes within a school or subject area which emphasize each factor, for example) are available from the Scoring Center at Northern Illinois University which services the CAQ.

Administration and Scoring: The CAQ can be administered by the classroom teacher in about 30 minutes. Printed on optical scanning forms for automated keypunching and computer scoring, test forms can be purchased and processed for $5.00 per class of 25. Completed questionnaires are sent to Dr. Robert Rosemier at Northern Illinois University, where they are scored, analyzed, and returned to individual teachers within 2 weeks. The CAQ's are usually treated by class, unless special analyses are requested. The user receives a 4-page computer printout containing: (1) frequency distributions and means of student responses to the open-ended items; (2) teacher intended and predicted responses for the same items; (3) results grouped by paired items with the proportion of individual students responding consistently calculated for each pair; (4) a summary of students' perceived emphases based on a scoring formula; and (5) a comparison of the students' perceptions with teacher intentions and predictions.

Comments: The Class Activities Questionnaire provides a measure of classroom climate by requiring students to make low-inference judgments about prevailing patterns of instruction. It is intended for use in large-scale evaluation of educational programs and in teacher self-assessment. CAQ authors have attempted to prevent its use as a tool for evaluating instructors, strongly discouraging administrators from using it "on" teachers. Considering its fairly recent development, this instrument is supported by substantial research, most of which confirms its validity.

References: Steele. J. M., *Class Activities Questionnaire (CAQ). General information.* Unpublished paper, 1969. (Available from the author, Evaluation Center, Mars Hill College, Mars Hill, North Carolina 28754).

Steele, J. M., *Availability and use of the Class Activities Questionnaire.* Unpublished paper, 1973. (Available from the author, Evaluation Center, Mars Hill College, Mars Hill, North Carolina 28754).

Steel, J. M., E. R. House, and T. Kerins, "An instrument for assessing instructional climate through low-inference student judgments." *American Educational Research Journal*, 1971, **8**, 447-466.

Wahlstrom, M. W., *Factorial validation of the Class Activities Questionnaire.* Paper presented at the meeting of the American Educational Research Association, New York, 1971.

Walberg, H. J., E. R. House, and J. M. Steele, "Grade level, cognition, and affect: A cross-section of classroom perceptions." *Journal of Educational Psychology*, 1973, **64**(2), 142–146.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Diagnostic Teacher-Rating Scale (#2)

AUTHOR: Sister Mary Amatora
AUTHOR/PUBLISHER ADDRESS: Cincinati, Ohio: Educators'-Employers' Tests and Services Associates, 1938

CRITERIA EVALUATED: 4, 7, 10, 11, 17, 18

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED
☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE (RETROSPECTIVE RATINGS)
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS
☐ AFFECTIVE
☐ COGNITIVE
☒ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☒ OTHER (STUDENT ATTITUDES)

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON (STUDENTS)
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR
☐ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 120-122.

Name: DIAGNOSTIC TEACHER-RATING SCALE

Author: Sister Mary Amatora

Type of Measure: Rating scale

Availability: Amatora, M., *Diagnostic teacher-rating scale.* Cincinnati, Ohio: Educators'-Employers' Tests and Services Associates, 1938.

Description: The Diagnostic Teacher-Rating Scale measures students' attitudes toward their teachers. The instrument was originally based on a list of teacher characteristics "liked" and "disliked" by a group of 200 children in grades 4 through 8. Three hundred teacher qualities were selected from this list. Items were written to represent these qualities and were organized on 7 scales: Liking for the Teacher; Ability to Explain; Kindliness, Friendliness, and Understanding; Fairness in Grading; Discipline; Amount of Work Required; and Liking for Lessons. Further revision reduced the number of items to 200, which were then sorted into 11 graduated categories, according to their estimated value. The 98 items determined most valuable were retained and arranged on 2 parallel forms (A and B), each containing 49 items, 7 on each scale. These 2 forms constitute the Diagnostic Check List, which is the major component of the Diagnostic Teacher-Rating Scale. The minor component is the Area Scale, a list of 7 questions corresponding to the 7 scales on the Check List.

Sample Items: The Area Scale contains questions which the student answers on a 5-point scale, ranging from "best" to "worst." The Diagnostic Check List, on the other hand, presents a series of qualities which may or may not apply to the teacher. The student simply checks those characteristics which he feels are typical of his teacher. The rating scale contains items of the following type:

Area Scale	*Best*				*Worst*
1. How clear are your teacher's explanations?	1	2	3	4	5
2. How much do you like your teacher?	1	2	3	4	5
3. How well does your teacher manage children in the classroom?	1	2	3	4	5

Diagnostic Check List

Scale I. Liking the Teacher

____1. Best teacher I've had.
____2. Smiles.
____3. Polite.
____4. Nice-looking.
____5. Doesn't keep the room neat.
____6. Doesn't have good sense of humor.
____7. Mean.

Reliability: The reliability of the Area Scale was investigated by Tschechtelin, Hipskind, and Remmers (1940). Using the split-half procedure on a group of 31 teachers and 610 students, they obtained coefficients ranging from .86 to .96.

The reliability of the Diagnostic Check List was estimated by correlating forms A and B. Data from a sample of 300 children produced coefficients ranging from .72 to .81 (reported in Buros, 1959).

Validity: Validity of the Diagnostic Teacher-Rating Scale can be approached in several ways (Remmers and Gage, 1955). First, since the instrument is designed to measure pupil attitudes, its validity may be assumed from its reliability. Or, if verbalized opinions are considered measures of attitude, the scale achieves construct validity to the extent to which it assesses verbalized opinions.

More concrete evidence of validity is provided by the scale's ability to differentiate between groups. A study by Tschechtelin, Hipskind, and Remmers (1940) indicates that scores discriminate between teachers selected by students as "best" and "poorest."

Correlations among the scales on the Diagnostic Check List and between the Area Scale and Diagnostic Check List also suggest validity. Correlations among the Check List scales are low (.06 to .33), while those between the Area Scale and the Check List are high (.74 to .90 for Form A, .75 to .91 for Form B).

Norms: Tentative norms are reported as percentile equivalents of the average ratings on each of the 7 scales.

Administration and Scoring: The Diagnostic Teacher-Rating Scale may be administered to students in grades 4 through 12 by principals, superintendents, supervisors, researchers, or teachers. When administered by the teacher being rated, suggestions are provided (Amatora, 1952) to insure honest responses from pupils. Though the scale is not timed, it usually takes about 30 minutes to complete.

The Area Scale is scored by adding the values of all responses to each of the 7 questions and then dividing each total by the number of pupils responding. Thus 7 scores are produced for the Area Scale. If a total score is desired, the 7 raw scores may be averaged.

The Diagnostic Check List is scored by computing the median number of items marked on each scale. The score for each scale is then found on a table. A total score for the Check List may be determined by finding the median value of these 7 scale scores.

The 2 parts of the instrument may be compared to find strengths, weaknesses, and areas needing improvement.

Comments: Although developed in 1938, the Diagnostic Teacher-Rating Scale remains a useful measure of students' attitudes toward the teacher, since item content is not at all dated. This scale has been frequently used in spite of

insufficient validity and normative data. Evidence of criterion-related validity, along with updated norms, would increase the value of this instrument.

References: Amatora, M., "Can elementary school children discriminate traits in their teachers?" *Child Development*, 1952, **23**, 75–79.

Amatora, M., "A diagnostic teacher-rating scale." *Journal of Psychology*, 1950, **80**, 395–399.

Amatora, M., "*Manual of directions for Diagnostic Teacher-Rating Scale.*" Cincinnati, Ohio: Educators'-Employers' Tests and Services Associates, 1952.

Buros, O. K., *The fifth mental measurements yearbook*. Highland Park, New Jersey: The Gryphon Press, 1959.

Remmers, H. H., and N. L. Gage, *Educational measurement and evaluation*. New York: Harper & Brothers, 1955.

Tschechtelin, M. A., M. J. F. Hipskind, and H. H. Remmers, "Measuring the attitudes of elementary-school children toward their teachers." *Journal of Educational Psychology*, 1940, **31**(3), 195–203.

Wrightsone, J. W., J. Justman, and I. Robbins, "*Evaluation in modern education.*" New York: American Book Co., 1956.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Illinois Teacher Evaluation Questionnaire (Adaptation for High School of the Illinois Course Evaluation Questionnaire) (#3)

AUTHOR: Lawrence M. Aleamone

AUTHOR/PUBLISHER ADDRESS: Measurement and Research Division, Office of Instructional Resources, University of Illinois, Urbana Illinois. (Also reported in Educational and Psychological Measurement, 1973, (33): 669-684).

CRITERIA EVALUATED: User must examine high school version

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY (AND COURSES)
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE (RETROSPECTIVE RATING)
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☒ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☒ OTHER (STUDENT ATTITUDES)

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON (STUDENTS)
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 123-125.

Name: THE ILLINOIS COURSE EVALUATION QUESTIONNAIRE

Author: Lawrence M. Aleamoni

Type of Measure: Questionnaire

Availability: Aleamoni, L. M., *The Illinois course evaluation questionnaire*. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972. A copy of Form 66 can also be found in: Aleamoni, L. M., and R. E. Spencer, "The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results." *Educational and Psychological Measurement*, 1973, **33**, 669–684.

Description: The Illinois Course Evaluation Questionnaire (CEQ) provides for student evaluation of courses and instructors. The original 1,000 items were based on criteria of effective instruction which were derived from an extensive review of the literature and suggestions from student and faculty committees at Pennsylvania State University. Following a trial administration of the instrument and factor analysis of the results obtained, 450 items were selected. Experimental forms of the test were readministered to 2 undergraduate samples at Pennsylvania State University. Results were analyzed, 28 items were retained, and 22 new items were added. This 50-item questionnaire (Form 66) was subjected to factor analysis, which produced 6 subscores.

A final revision of the CEQ, Form 73, was developed in 1973 to eliminate redundancy in the first form. It contains 23 items and yields 5 subscores: General Course Attitude, Course Content, Method of Instruction, Interest and Attention of the Students, and the Instructor. In addition, the teacher may include optional items which he has devised or selected from a catalogue of supplementary items.

The CEQ was designed for use with college-level students, although its content appears appropriate for high school seniors and juniors. A variation of the scale, the Illinois Teacher Evaluation Questionnaire (Mahan, 1970), has been constructed specifically for use in high school classrooms, but limited psychometric information is available on this form.

The CEQ can be used to provide information to both administrators and instructors in an effort to improve the quality of teaching.

Sample Items: The 23 items of Form 73 of the CEQ are followed by 4 response alternatives: Strongly Agree, Agree, Disagree, and Strongly Disagree. The instrument contains items of the following type:

1. I do not want to take another course from this instructor.
2. I would have preferred a different method of teaching in this course.
3. The instructor seemed to have an excellent knowledge of the subject matter.

Reliability: The reliability of the CEQ (Form 66) has been investigated in several different studies. Using split-half methods, 2 of these studies reported reliabilities of .92 and .93, corrected for length by the Spearman-Brown formula. The first study correlated negative and positive items, while the second correlated both halves of the negative items and both halves of positive items (reported in

Aleamoni, 1972). Reliabilities for each of the subscore areas were computed using the Kuder-Richardson formula 21 on the responses of students from 7 different courses. These ranged from .40 to .92, with Content being the least reliable subscale (Aleamoni and Spencer, 1973).

Gillmore (1972) also conducted several reliability studies. On a sample of 5,346 students, he obtained CEQ item reliabilities ranging from .73 to .94; subscale reliabilities ranged from .80 to .98. In a second study, using ratings of 103 instructors teaching 2 sections of the same course, he obtained item reliabilities ranging from .53 to .82. Dividing these sections into 2 groups and correlating odd and even items within each group, Gillmore found that coefficients ranged from .54 to .91 for Group I and from .67 to .91 for Group II.

Validity: Many different studies have investigated the validity of the CEQ (Form 66). Spencer and Dick (1965) related CEQ subscores to sex, term, curriculum, and final grade. Only final grade was significantly correlated with CEQ scores, a finding which is consistent with previous research.

Stallings and Spencer (1967) asked 10 measurement specialists and teaching assistants from the Speech Department of the University of Illinois to rate nine instructors. Their evaluations were correlated with student ratings, and a significant coefficient of .70 was obtained.

In a study by Swanson and Sisson (1971), a rank-order correlation of .70 was found between peer and student ratings of 12 instructors at Bowling Green State University. Aleamoni and Yimer (1972) found that colleague and student ratings on the CEQ were not significantly related to an instructor's research productivity. Yet, colleague ratings were significantly and positively related to an instructor's academic rank. This suggests that an instructor's reputation may also influence the ratings given him.

Finally, Aleamoni and Hexner (1973) found that students gave instructors higher ratings when they were told that their evaluations would be used in making salary and promotion decisions. These results suggest that information given preceding a course evaluation is an important variable to consider when administering the CEQ.

Norms: Normative data have been derived from the ratings of approximately 750 courses containing 2,784 sections and over 100,000 students at the University of Illinois. Additional normative information is provided on 5,346 course sections taught throughout the United States. These norms are continually updated.

Administration and Scoring: The CEQ is usually administered in the classroom over a 10-minute period. It is machine scored, and results are returned to the instructor. Average subscores and total scores with norm deciles are provided, along with specific item responses and their means and norm deciles. Pooled results for a number of sections of the same course may also be obtained.

Comments: The Illinois Course Evaluation Questionnaire is a well constructed instrument with adequate reliability, validity, and normative data provided for Form 66. The CEQ is an economical means of supplying feedback to individual instructors or evaluating the teaching process on a larger scale. The high school

version, The Illinois Teacher Evaluation Questionnaire, should become equally useful with the accumulation of relevant psychometric data.

References: Aleamoni, L. M., *A review of recent reliability and validity studies of the Illinois Course Evaluation Questionnaire* (*CEQ*). Research Memorandum No. 127. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972.

Aleamoni, L. M., *Evaluation by students to identify general instructional problems.* Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1973.

Aleamoni, L. M., *The evaluation of instruction and the use of CEQ, Form 73.* Research Memorandum No. 152. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1974.

Aleamoni, L. M., and P. Z. Hexner, *The effect of different sets of instructions on student course and instructor evaluation.* Research Report 339. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1973.

Aleamoni, L. M., and R. E. Spencer, "The Illinois Course Evaluation Questionnaire: A description of its development and a report of some of its results." *Educational and Psychological Measurement*, 1973, **33**, 669–684.

Aleamoni, L. M., and M. Yimer, "An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness." *Journal of Educational Psychology*, 1973, **64**(3), 274–277.

Gillmore, G. M., *Estimates of reliability coefficients for items and subscales of the Illinois Course Evaluation Questionnaire.* Report No. 341. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1972.

Mahan, J. M., *Manual of interpretation for the Illinois Teacher Evaluation Questionnaire.* Research Report No. 316. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1970.

Spencer, R. E., and W. Dick, *Course evaluation questionnaire: Manual of interpretation.* Research Report No. 200. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1965.

Stallings, W. M., and R. E. Spencer, *Ratings of instructors in Accountancy 101 from video-tape clips.* Research Report No. 265. Urbana, Illinois: Measurement and Research Division, Office of Instructional Resources, University of Illinois, 1967.

Swanson, R. A., and D. J. Sisson, "The development, evaluation, and utilization of a departmental faculty appraisal system." *Journal of Industrial Teacher Education*, 1971, **1**, 64–79.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Learning Environment Inventory (LEI) (LEI #4)

AUTHOR: Gary J. Anderson
AUTHOR/PUBLISHER ADDRESS: Atlantic Institute of Education, Halifax, NS

CRITERIA EVALUATED: 6, 11, 12, 14, 15 (Indirect, based on retrospective ratings by students)

SUBJECT OF OBSERVATION

____ TEACHER AND PUPILS
____ TEACHER ONLY
____ PUPIL ONLY
__x__ SMALL GROUPS (CLASSROOM CLIMATE)

SETTINGS

__x__ CLASSROOM, ANY SUBJECT (GR 7-12)
____ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

____ CATEGORY CHANGE
____ TOPIC CHANGE
____ SPEAKER CHANGE
____ TIME UNIT
____ TIME SAMPLE
__x__ RATING SCALE
____ OTHER

NUMBER OF SUBJECTS OBSERVED

____ 1 ONLY
__x__ MORE THAN 2 PEOPLE IN CLASSROOM (ENTIRE CLASS)

COLLECTION METHODS

__x__ LIVE (RETROSPECTIVE RATING)
____ LIVE WITH CODING EQUIPMENT
____ VIDEO AND/OR AUDIO TAPE

OBSERVER(S) REQUIRED

__x__ 1 PERSON (STUDENTS)
____ TEAM OF 2 PERSONS
____ 2 TEAMS OF 2 PERSONS

CATEGORY DIMENSIONS

__x__ AFFECTIVE
____ COGNITIVE
____ PROCEDURE OR ROUTINE
____ ACTIVITY
__x__ SOCIOLOGICAL STRUCTURE
____ OTHER

USES REPORTED BY AUTHOR

____ RESEARCH
____ TRAINING
__x__ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 412-143.

Name: LEARNING ENVIRONMENT INVENTORY

Author: Gary J. Anderson

Type of Measure: Rating scale

Availability: Anderson, G. J., *The assessment of learning environments: A manual for the Learning Environment Inventory and the My Class Inventory.* Halifax, Nova Scotia, Canada: Atlantic Institute of Education, 1973.

Description: The Learning Environment Inventory (LEI) measures classroom climate as perceived by secondary school pupils along 15 dimensions, reflecting both interpersonal relationships among students and organizational characteristics of the classroom.

The LEI was developed in an attempt to expand and improve Walberg's (1968) Classroom Climate Questionnaire. Seven items were included on each LEI scale to obtain internally consistent dimensions. Items were assigned to scales by 4 independent judges. Those which were misclassified by one or more of the judges were modified or replaced. After empirical testing, 6 items with poor scale correlations were revised. The final version of the LEI contains 105 statements describing typical school classes on the following scales: Cohesiveness, Diversity, Formality, Speed, Environment, Friction, Goal Direction, Favoritism, Cliqueness, Satisfaction, Disorganization, Difficulty, Apathy, Democratic, and Competitiveness.

This instrument can be used to assess an individual student's perception of his class or to measure the learning environment of the class as a whole. Anderson (1973) maintains that class means provide the best estimate of collective student perceptions and suggests their use when different treatments are being studied across classes.

Sample Items: The LEI contains items of the following type with which the respondent expresses his agreement or disagreement on a 4-point scale.

Scale	*Item*
Cohesiveness	1. Students in this class don't know each other well enough to like or dislike one another.
Formality	2. There are well established rules of behavior in this classroom.
Friction	3. There is an atmosphere of tension among groups of students in this class.
Satisfaction	4. Most students enjoy coming to class.

Reliability: Alpha coefficients, intraclass correlations, and test/retest reliabilities are reported in the LEI manual (Anderson, 1973). Alpha coefficients, based on the responses of 1,048 high school students range from .54 (Diversity) to .85 (Goal Direction), while intraclass correlations for the same sample fall between .31 (Diversity) and .92 (Formality and Disorganization). Test/retest estimates, computed on a sample of 139 11th- and 12th-graders, over a 4-week interval. range from .43 (Diversity) to .73 (Friction). Anderson, notes, however, that

these results may be unrepresentative since unusual national political events, believed to have had a strong effect on high school students, occurred between test administrations.

Validity: The LEI manual summarizes validity studies on a scale-by-scale basis. Major relationships between each scale and selected other variables are reported. For example, the Cohesiveness scale has been related to class size, with smaller groups indicating greater cohesiveness. Large classes, on the other hand, were rated as more formal than small classes.

Science classes using the Harvard Project Physics Course, which promotes a variety of interests and activities and attempts to simplify physics, received higher scores on the Diversity scale and lower scores on the Difficulty scale than more traditional physics classes.

Other relationships, too numerous to mention here, were also in the expected directions.

Norms: Scale means and standard deviations are provided for 1,048 high school students in a variety of subject areas. Class means are also presented for 2 samples: (1) 47 Harvard Project Physics classes and 37 traditional physics classes, whose students were given an early (1967) version of the LEI; and (2) 62 diverse classes from 8 high schools in Montreal, where the 1969 version was administered.

Administration and Scoring: The test can be easily administered in 30 to 40 minutes. Students can respond on an optical scanning answer sheet, and their responses can then be punched onto regular IBM computer cards. Machine scoring is recommended due to the large number of subscales. Response alternatives are assigned values from 1 to 4. Scale scores are computed by totaling the item scores for each scale. Since all scales contain 7 items, scale scores range from 7 to 28.

Comments: The LEI is short, easy to administer, and supported by a wide variety of research. In general, its phychometric characteristics are satisfactory. The moderate internal consistency reliabilities are acceptable since the LEI attempts to measure many dimensions validly, rather than to measure a few dimensions very reliably.

References: Anderson, G. J., *The assessment of learning environments: A manual for the Learning Environment Inventory and the My Class Inventory.* Halifax, Nova Scotia, Canada: Atlantic Institute of Education, 1973.

Edwards, K. J., D. L. DeVries, and S. A. Livingston, *Changing the focus of response in assessing classroom learning environments.* Baltimore: Center for Social Organization of Schools, 1973.

Walberg, H. J., "Social environment as a mediator of classroom learning." *Journal of Educational Psychology,* 1969, **60**, 443–448.

Walberg, H. J., and G. J. Anderson, "Properties of the achieving urban classes." *Journal of Educational Psychology,* 1972, **63**, 381–385.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Pupil Observation Survey Report (#5)

AUTHOR: Donald J. Veldman and Robert F. Peck

AUTHOR/PUBLISHER ADDRESS: Research and Development Center for Teacher Education, The University of Texas, Austin, Texas, 78712. (Copy in Journal of Educational Psychology, 1963 (54): 346-355).

CRITERIA EVALUATED: 4, 7, 10

SUBJECT OF OBSERVATION

TEACHER AND PUPILS
x TEACHER ONLY
PUPIL ONLY
SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

x 1 ONLY
MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

x LIVE
LIVE WITH CODING EQUIPMENT
VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

x AFFECTIVE
x COGNITIVE
x PROCEDURE OR ROUTINE
x ACTIVITY
SOCIOLOGICAL STRUCTURE
x OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

CATEGORY CHANGE
TOPIC CHANGE

SPEAKER CHANGE
TIME UNIT
TIME SAMPLE
x RATING SCALE
OTHER

OBSERVER(S) REQUIRED

x 1 PERSON
TEAM OF 2 PERSONS
2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

x RESEARCH
x TRAINING
EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 128-130.

Name: PUPIL OBSERVATION SURVEY REPORT

Authors: Donald J. Veldman and Robert F. Peck

Type of Measure: Rating scale

Availability: The Pupil Observation Survey Report can be obtained from the Research and Development Center for Teacher Education, The University of Texas, Austin, Texas 78712. A copy of this instrument also appears in: Veldman, D. J., and R. F. Peck, "Student teacher characteristics from the pupils' viewpoint." *Journal of Educational Psychology*, 1963, **54**, 346–355; and in White, W. F., and O. T. Dekle, "Effect of teacher's motivational cues on achievement level in elementary grades." *Psychological Reports*, 1966, **18**, 351–356.

Description: The Pupil Observation Survey Report (POSR) is a 38-item rating scale used to measure pupils' perceptions of their student teacher. Though originally designed for use in grades 7 through 12, the POSR has been successfully employed by White and Dekle (1966) in grades 5 and 6 as well.

Items on the POSR represent 5 factors: (1) Friendly, Cheerful, Admired; (2) Knowledgeable, Poised; (3) Interesting, Preferred; (4) Strict Control; and (5) Democratic Procedure. Factor 1 has been shown to differentiate ($p < .01$) overachievers, underachievers, and average achievers (White and Dekle, 1966). On Factors, 2, 4, and 5, cooperating teachers received higher ratings than student teachers, while the reverse was true on Factors 1 and 3 (Veldman, 1970).

Sample Items: All of the items on the POSR are positively phrased statements describing the student teacher. The pupils respond to these statements in 1 of 4 ways: "T" = completely true; "t" = more true than false; "f" = more false than true; and "F" = totally false. Items of the following type appear on the scale:

Factor	*Item*
1. Friendly, Cheerful, Admired	She is very easy to get along with.
2. Knowledgeable, Poised	She knows a lot about the subject she teaches.
3. Interesting, Preferred	With her, learning is more fun than it is work.
4. Strict Control	Her students can't get away with anything.
5. Democratic Procedure	She often asks the students how they feel about proposed projects before assigning them.

Reliability: Veldman and Peck (1963) report factor score reliabilities for the POSR. These were computed on 50 student teachers (selected from a sample of 554) whose classes were the largest. Each class was randomly divided in half, and each of these halves completed the POSR. Factor scores were obtained for each half-class. The two comparable factor scores for each teacher were then

correlated to yield "split-class" reliabilities of: .92 for Factor 1; .72 for Factor 2; .91 for Factor 3; .81 for Factor 4; and .89 for Factor 5. These estimates were not corrected for size since they were computed on the largest classes in the sample.

Validity: Veldman and Peck (1963) approached construct validity by checking the stability of POSR factors across different samples. They compared factor structures based on the scores of three separate groups of student teachers and obtained coefficients ranging from .89 to .99 (median = .98).

In an attempt to establish criterion-related validity, POSR factor scores were correlated with scale scores on the California Personality Inventory (CPI) and the Self-Report Inventory (SRI). Correlations between POSR and CPI scores never exceeded .30, although 34 out of 180 were statistically significant. Those between the POSR and the SRI did not exceed .24. These correlations indicated minor, but indistinct, relationships between the measures. Correlations between the POSR scores and supervisor ratings of student teachers yielded more interpretable results. Factors 1, 2, and 4 (Friendliness, Knowledge, and Control) were all significantly related to the supervisory ratings of teaching effectiveness (Veldman and Peck, 1963).

In a later study, Veldman and Peck (1969) investigated the effect of content and other variables on POSR scores. Their data clearly indicated that the subject matter being taught strongly affected POSR scores, while grade level and socioeconomic status of the respondents exerted only minor influence on ratings given to teachers.

Norms: Veldman and Peck (1963) present no normative data for the POSR. However, use of local norms is recommended for this instrument.

Administration and Scoring: Administering the POSR is fairly easy. Pupils are given a set of instructions and an explanation of the test, assuring them that their responses will not be used to give the student teacher a grade. It is preferable that the POSR be answered anonymously and administered by someone other than the student teacher.

Scores on the POSR are determined by assigning numerical values (1 through 4) to the various response categories. Since all items are positively stated, scoring is never reversed. Factor scores are computed by summing the scores of items representing each factor.

Comments: As a research instrument, the Pupil Observation Survey Report has considerable utility. It can also be used in teacher training to help student teachers assess their performances and identify problems in pupil/teacher relationships. Users of the POSR, however, must recall that subject matter being taught may affect POSR scores, causing certain groups of student teachers to obtain consistently lower ratings.

References: Veldman, D. J., "Pupil evaluation of student teachers and their supervisors." *Journal of Teacher Education*, 1970, **21**, 165–167.

Veldman, D. J., and R. F. Peck, "Student teacher characteristics from the pupils' viewpoint." *Journal of Educational Psychology*, 1963, **54**, 346–355.

Veldman, D. J., and R. F. Peck, "Influences on pupil evaluations of student teachers." *Journal of Educational Psychology*, 1969, **60**, 103–108.

White, W. F., and O. T. Dekle, "Effects of teacher motivational cues on achievement level in elementary grades." *Psychological Reports*, 1966, **18**, 351–356.

OBSERVATION INSTRUMENT DESCRIPTION

NSTRUMENT: The Purdue Teacher Evaluation Scale (#6)

UTHOR: Ralph R. Bentley and Allan R. Starry

UTHOR/PUBLISHER ADDRESS: Purdue Research Foundation, Purdue University, Lafayette, Indiana.

RITERIA EVALUATED: 7, 12, 17

UBJECT OF OBSERVATION

x TEACHER AND PUPILS
___ TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

UMBER OF SUBJECTS OBSERVED

x 1 ONLY
___ MORE THAN 2 PEOPLE IN CLASSROOM

OLLECTION METHODS

x LIVE (RETROSPECTIVE RATING)
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

ATEGORY DIMENSIONS

x AFFECTIVE
___ COGNITIVE
x PROCEDURE OR ROUTINE
___ ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

___ CATEGORY CHANGE
___ TOPIC CHANGE
___ SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
x RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON (STUDENTS)
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

___ RESEARCH
___ TRAINING
x EVALUATION*

EFERENCE:

valuating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan . Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 134-35.

* INDIVIDUAL ITEM INFORMATION MORE USEFUL THAN SCALE SCORES

Name: THE PURDUE TEACHER EVALUATION SCALE

Authors: Ralph R. Bentley and Allan R. Starry

Type of Measure: Rating scale

Availability: Bentley, R. R., and A. R. Starry, *The Purdue teacher evaluation scale, form A*. Lafayette, Indiana: Purdue Research Foundation, Purdue University, 1970. *The Purdue Teacher Evaluation Scale* and the *Manual for the Purdue Teacher Evaluation Scale* may be obtained from the University Book Store, 360 State Street, West Lafayette, Indiana 47906.

Description: The Purdue Teacher Evaluation Scale (PTES) was designed to provide the junior or senior high school teacher with an evaluation of his performance as seen through the eyes of his students. The PTES contains 60 questions which are grouped into 6 scales identified by a composite principal components analysis. These scales are: (1) Ability to Motivate Students; (2) Ability to Control Students; (3) Subject Matter Orientation of Teacher; (4) Student-Teacher Communication; (5) Teaching Methods and Procedures; and (6) Teacher Fairness.

The PTES was developed to aid teachers in their efforts at self-improvement. Accordingly, students are asked to assess specific, rather than general, behavioral characteristics. Scale items were originally drawn from a collection of short statements about teacher behaviors, written by educators and researchers. These statements were administered to approximately 60 classes of students, and were then analyzed and revised to make up the present form of the PTES.

Sample Items: Students respond to PTES items by selecting and checking 1 of 4 alternatives, ranging from "very much like my teacher" to "very much unlike my teacher." Items of the following type are included on the scale.

My teacher . . .

1. speaks clearly.
2. knows when we do not understand.
3. has pet students.
4. gives homework which helps in understanding the lessons.

Reliability: Bentley and Starry (1970) maintain that test/retest reliabilities are most meaningful for a rating scale of this type. However, since test/retest data are not available, they report an estimate of stability obtained using a modified split-half technique. Twenty evaluations for each of 128 teachers were selected and divided into random halves. For each random half, a median score for each scale was computed. These 128 median scores were then correlated, producing reliability coefficients ranging from .83 for Teacher Fairness to .91 for Ability to Motivate Students.

Validity: Bently and Starry (1970) report that there are no relevant criteria by

which to judge the validity of the PTES. They do note, however, that students tend to agree with one another in rating a particular teacher.

Correlations among the six PTES scales, computed on a sample of 198 teachers, range from .47 to .89.

Norms: Norms are based on a sample of 198 teachers, evaluated on the PTES by approximately 5,000 students. These teachers were from middle- and upper-class junior and senior high schools located in the midwestern United States.

Administration and Scoring: The PTES is self-administering, though the instructions, printed on a separate sheet, must be read to the students, preferably by someone other than the teacher being evaluated. Approximately 20 minutes are required to complete the PTES.

Scoring may be done either by hand or by computer. Sixty median item scores and 6 scale scores are produced for each class. For positive statements the response "Very much like my teacher" is weighted 4 and the response "Very much unlike my teacher" is weighted 1. For negatively worded items the weights are reversed. While a scoring key is provided, a more comprehensive scoring and reporting service is offered by Purdue University's Measurement and Research Center. If the teacher desires, PTES results are treated confidentially.

Comments: The Purdue Teacher Evaluation Scale is an excellent aid for the teacher interested in improving his performance. Though the value of scale scores is questionable (due to high scale intercorrelations), the PTES does provide valuable information via 60 item scores. It also offers adequate normative data (a rarity among teacher evaluation questionnaires) and a confidential scoring service.

References: Bentley, R. R., and A. R. Starry, *Manual for the Purdue Teacher Evaluation Scale.* Lafayette, Indiana: Purdue Research Foundation, Purdue University, 1970.

Purdue measurement and evaluation instruments. Lafayette, Indiana: Purdue Research Foundation, Purdue University, undated.

Woolfold, T. V., *Teacher perception of organizational inducement and student assessment of teacher performance.* Unpublished doctoral dissertation, Purdue University, 1970.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Students' Evaluations of Instructional Effectiveness: Research and Survey Instrument (#7).

AUTHOR:
AUTHOR/PUBLISHER ADDRESS: Beatty, Brondon; Marsh, Herbert W. (ERIC Microfiche, ED 141 413)

CRITERIA EVALUATED: 1, 6, 10, 11, 12, 15

SUBJECT OF OBSERVATION

☒ TEACHER AND PUPILS
☐ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☐ 1 ONLY
☒ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE (RETROSPECTIVE RATING)
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☒ PROCEDURE OR ROUTINE
☒ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE
☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON (SENIOR STUDENTS)
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

ERIC ED 141 413

TABLE I

DIMENSIONS OF EFFECTIVE TEACHING AND THEIR DEFINING ITEMS

BREADTH

30 -- The instructor discussed points of view other than his own
47 -- The instructor contrasts the implications of various theories
48 -- The instructor adequately discusses recent developments in the field
49 -- The instructor adequately presents the origin of ideas and concepts
56 -- Does the instructor invite criticism of his own ideas

ORGANIZATION/CLARITY

51 -- Relative to other courses, the instructor's lectures are easy to outline
2 -- The course objectives were adequately outlined so that you knew where the course was going
35 -- Adequate time was distributed over course topics
1 -- The purpose of class assignments was made clear and was integrated into the overall course structure
28 -- The instructor was helpful in emphasis and summarization of main points during presentations
50 -- The instructor's explanations are clear
24 -- The course successfully integrated lecture materials and outside assignments
35 -- To what extent was the workload evenly distributed over the quarter
23 -- Agreement between proposed course objectives and those actually taught

INTEREST/DYNAMIC

61 -- The instructor is dynamic and energetic
62 -- The instructor has an interesting style of presentation
63 -- The instructor enjoys teaching
5 -- The instructor was enthusiastic about teaching the course
18 -- What is your overall rating of the instructor
64 -- The instructor has a sense of humor
32 -- The instruct held your interest

DIFFICULTY

38 -- Did the instructor pace his presentations to the understanding of the class
17 -- Relative to other courses, how difficult was it to master the contents of the course
34 -- The instructor accurately gauged student background and knowledge in the subject

LEARNING/VALUE

15 -- You learned something of value from the course
46 -- You have developed an increased appreciation for the subject
16 -- How adequately do you feel that you have mastered the contents of the course
40 -- The examinations were of instructional value
41 -- The assignments were interesting and stimulating
20 -- To what extent did the course meet your objectives for taking the course
13 -- The instructor stimulated your intellectual curiosity in the subject

INTERACTION/DISCUSSION

53 -- Does the instructor encourage class discussion
8 -- Given the constraints of the course, does the instructor involve students in class discussion
54 -- Does the instructor adequately invite students to share their knowledge and experiences
26 -- Does the instructor allow a few students to monopolize class discussion
59 -- The instructor relates to students as individuals

INDIVIDUAL CONTACT

9 -- The instructor had adquate office hours during which he/she was available
60 -- The instructor recognizes and greets students out of class
58 -- The instructor is friendly toward students
10 -- The instructor was easy to approach and made you feel welcome in seeking help
57 -- The instructor has a genuine interest in the students
59 -- The instructor relates to students as individuals

OBSERVATION INSTRUMENT DESCRIPTION

NSTRUMENT: Student Evaluation of Teaching I (#8)

UTHOR: Donald J. Velma and Robert F. Peck

UTHOR/PUBLISHER ADDRESS: Dissemination Dept., R and D Center for Teacher Education, The University of Texas, Austin, Texas 78712

RITERIA EVALUATED: 7, 17

UBJECT OF OBSERVATION

___ TEACHER AND PUPILS
x TEACHER ONLY
___ PUPIL ONLY
___ SMALL GROUPS

UMBER OF SUBJECTS OBSERVED

x 1 ONLY
___ MORE THAN 2 PEOPLE IN CLASSROOM

OLLECTION METHODS

x LIVE (RETROSPECTIVE RATING)
___ LIVE WITH CODING EQUIPMENT
___ VIDEO AND/OR AUDIO TAPE

ATEGORY DIMENSIONS

x AFFECTIVE
___ COGNITIVE
x PROCEDURE OR ROUTINE
___ ACTIVITY
___ SOCIOLOGICAL STRUCTURE
___ OTHER

SETTINGS

x CLASSROOM, ANY SUBJECT
___ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

___ CATEGORY CHANGE
___ TOPIC CHANGE

___ SPEAKER CHANGE
___ TIME UNIT
___ TIME SAMPLE
x RATING SCALE
___ OTHER

OBSERVER(S) REQUIRED

x 1 PERSON (STUDENTS)
___ TEAM OF 2 PERSONS
___ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

___ RESEARCH
___ TRAINING
x EVALUATION

EFERENCE:

valuating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan . Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 138-39.

Name: STUDENT EVALUATION OF TEACHING I

Authors: Donald J. Veldman and Robert F. Peck

Type of Measure: Rating scale

Availability: Requests should be sent to SET I Order, Dissemination Department, Research and Development Center for Teacher Education, The University of Texas, Austin, Texas 78712.

Description: The Student Evaluation of Teaching (SET) I is a 10-item scale designed to measure the way in which students in fifth grade or above perceive their teachers. The SET I was derived from a factor analysis of the Pupil Observation Survey Report (POSR) (Veldman and Peck, 1967), which yielded 5 major dimensions describing student perceptions of teacher style. These were: Friendly and Cheerful; Knowledgeable and Poised; Lively and Interesting; Firm Control (Discipline); and Nondirective (Democratic). The two items loading highest on each factor were selected for inclusion on the SET I. Data used in developing the SET I were collected only on the POSR and then applied to SET I items.

The authors suggest that this instrument be used in evaluating student teachers, as a source of information concerning pupil/teacher interaction. A companion instrument, the SET II (included in this volume), may be used to evaluate kindergarten through third-grade teachers.

Sample Items: The 10 items on the SET I are divided into 5 scales, each containing 2 items. The student responds to each item by marking F for "always false," f for "often false," t for "often true," and T for "always true," depending on the degree to which the statement applies to the teacher in question.

The following items are similar to those on the SET I:

This teacher:

1. F f t T is always kind to students.
2. F f t T is very interesting.
3. F f t T does not let students misbehave.

Reliability: Since the research report describing the SET I presents no data collected specifically on the SET I, no reliability figures are available. Veldman (1970) does report intercorrelations between the 2 items on each scale, ranging from .56 to .86. These data were obtained by deriving SET scores from POSR data.

Validity: Validity studies on the SET I are also scarce. However, the items clearly relate to the student's perception of the teacher. Therefore, the instrument seems to have face validity.

A factor analytic study reported by Veldman (1970) attempted to confirm that the SET I factor structure corresponded to that of the POSR. Each item

loaded highest on the expected factor. Unfortunately, the SET I data used in this study were simulated from the POSR data to which they were compared.

Norms: No norms are provided for the SET I. The nature and purposes of this instrument preclude normative expectations. Veldman and Peck (1970) report several tendencies in the way in which particular kinds of teachers are perceived. These should be considered when one is interpreting SET I results. Variables that are apparently related to SET I scores are teacher and pupil sex, subject matter area, grade, and socioeconomic level.

Administration and Scoring: Administration of the SET I is not explicitly discussed in the research report describing the instrument. However, it appears that administration is quite easy. Instructions are printed on the test sheets, which may be given to an entire class of pupils at once.

Scoring the SET I is also fairly easy. All items are stated in a positive direction: F = 1; f = 2; t = 3; and T = 4. A computer is useful in processing group data. Student responses can be punched onto computer cards according to a program recommended by Veldman (1970), which produces item means and scale scores, giving a verbal interpretation of both. Finally, the average of all item means is computed as an index of overall evaluation. All of this can be done by hand, but it would be a laborious procedure.

Veldman (1970) emphasizes the importance of interpreting scale scores as complete profiles. For example, a high score on Firm Control would mean different things depending on the level of other scale scores.

Comments: Though the SET I lacks adequate psychometric data, it is probably an appropriate instrument for the purposes suggested by its authors. It does yield a summary measure of students' perceptions of a teacher. Such an assessment can provide useful information to the teacher concerning his or her impact on students. Users of the SET I, however, should be careful not to over-interpret the results, since so few psychometric data are available.

References: Veldman, D. J., *Student evaluation of teaching*. Research Methodology Monograph No. 10. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1970.

Veldman, D. J., and R. F. Peck, *Information about Student Evaluation of Teaching*. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, undated.

Veldman, D. J., and R. F. Peck, *The Pupil Observation Survey: Teacher behavior from the pupils' viewpoint*. Austin, Texas: Research and Development Center for Teacher Education, University of Texas, 1967.

OBSERVATION INSTRUMENT DESCRIPTION

ISTRUMENT: Student Evaluation of Teacher Instrument II (#9)

UTHOR: Ruth Adolph Haak, Douglas A. Kleiber, and Robert F. Peck

UTHOR/PUBLISHER ADDRESS: Dissemination Dept., R and D Center for Teacher Education, University of Texas, Austin, Texas, 78712

RITERIA EVALUATED: 10, 17

UBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

UMBER OF SUBJECTS OBSERVED

☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

OLLECTION METHODS

☒ LIVE (RETROSPECTIVE RATINGS)
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

ATEGORY DIMENSIONS

☒ AFFECTIVE
☐ COGNITIVE
☒ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☐ OTHER

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON (STUDENTS)
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☐ TRAINING
☒ EVALUATION

EFERENCE:

valuating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan . Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 140-42.

Name: STUDENT EVALUATION OF TEACHER INSTRUMENT II

Authors: Ruth Adolf Haak, Douglas A. Kleiber, and Robert F. Peck

Type of Measure: Rating scale

Availability: Requests should be sent to: SET II Order, Dissemination Department, Research and Development Center for Teacher Education, University of Texas, Austin, Texas 78712.

Description: The Student Evaluation of Teacher Instrument II (SET II) contains 20 items, each of which is a simple statement about the teacher or the teacher's attitude toward the student. It incorporates 3 dimensions, derived through factor analysis of an early form of the instrument. These are: Unreasonable Negativity; Fosterance of Self-Esteem; and Stimulating, Interactive Style, which has two subscales, Rapport and Interactional Competence.

The SET II was designed to assess the way in which children in kindergarten through third grade (or disadvantaged children through grade 6) view their teachers. Its predecessor, the SET I (included in this volume), is used with older children, in grades 5 and above. In developing the initial form of SET II, the authors attempted to construct items that would measure either students' perceptions of teacher competency or their feeling ("love") toward the teacher. Items were originally classified as either direct assessments of the teacher or estimates of the teacher's attitude toward students. Under each classification were the dimensions of "love," "competency," and "diagnostic information." Factor analysis of the pilot data failed to substantially support these classifications, and they were abandoned in favor of the factor structure mentioned above.

Sample Items: The SET II contains items of the following type:

Factor	*Item*
I. Stimulating, Interactive Style	
Subscale 1: Rapport (5 items)	1. My teacher makes school a lot of fun. 2. My teacher thinks kids are good.
Subscale 2: Interactional Competence (4 items)	1. My teacher listens to what we say.
Factor	*Item*
II. Unreasonable Negativity (5 items)	1. My teacher thinks I am lazy. 2. My teacher is always "picking on" me.
III. Fosterance of Self-Esteem (6 items)	1. My teacher thinks I can do a lot by myself. 2. My teacher thinks I have good ideas.

Each item is printed on a separate card and is read to the student who then places the card in either a "true" or "false" pile.

Reliability: Haak, Kleiber, and Peck (1972) report test/retest reliability data collected during the pilot study of the SET II. Since item format calls for a true-false answer, the authors chose a percentage of agreement (PA) statistic as a measure of reliability. The sample was taken from children in kindergarten through grade 6 in the Austin (Texas) Independent School District. An attempt was made to choose a stratified sample in terms of grade level and cultural background. Percentages of agreement among children in kindergarten ranged from 59.3 to 89.5, with 8 items achieving PAs greater than 75. PAs for first-grade children ranged from 69.7 to 92.2, with 15 being greater than 80. The range for second- and third-grade children was 68.3 to 89.4, with 16 PAs at 75 or above. An additional sample composed of fourth- through sixth-grade Mexican-American children yielded PAs ranging from 66.9 to 94, with 19 over 75. The authors also report percentage of agreement from first to second testing on a sample divided in terms of cultural group (black, Mexican-American, and white), ability level (low, medium, and high), and a combination of cultural group and ability level. Except at the kindergarten level, PAs for each item were generally between 70 and 90.

Validity: No validity data are reported in the SET II manual (Haak, Kleiber, and Peck, 1972). However, factor analytic data for grades 1 through 3 in the pilot study are presented. Factor loadings are fairly high: those on Factor 1 ranged from .60 to .94; those on Factor 2 from .61 to .81; and those on Factor 3 from .59 to .78. Most items did not load substantially in the same direction on more than 1 factor, the only exception being item 19 ("She likes for me to help her"), which loaded −.45 on Factor 2 and −.59 on Factor 3. A factor analysis of data collected from all elementary grades produced similar factor loadings. Items on Factor 1, however, split into 2 groups: Rapport and Interactional Competence. These 2 item groupings were incorporated as subscales of Factor 1.

Norms: The norms presented in the SET II manual are based on only 36 teachers in the public schools of Austin, Texas. The SET II authors do not recommend their use in other settings; instead they encourage local norming.

Administration and Scoring: The procedure for administering the SET II is uncomplicated and requires no special training. On each card an item is printed, along with a picture which is placed in the upper right-hand corner. The administrator identifies the card to the children by calling attention to the picture. He or she then reads the item to the children, who respond to it by placing the card in either a "true" envelope, identified by a picture of a post-office box, or a "false" envelope, identified by a picture of a trash can.

A tally sheet is used to score each student's responses. The scorer checks "true" or "false" beside each item listed. Responses are weighted as indicated on the scoring sheet, and then summed to obtain a scale score.

Teacher tally sheets and profile sheets may be used to record the entire class' perception of the teacher.

Comments: The SET II is a carefully constructed instrument with a great deal of face validity. However, there are as yet no real validity or reliability studies relating to scale or subscale scores. With more developmental research in these 2 areas, the SET II may become a useful instrument for helping teachers in grades 1 through 3 to develop interpersonal skills and more positive relationships with their pupils.

References: Haak, R., D. Kleiber, and R. Peck, *Manual: Student Evaluation of Teacher Instrument, II*. Austin, Texas: The Research and Development Center for Teacher Education, The University of Texas, 1972.

Borich, G. D., R. C. Godbout, R. F. Peck, M. M. Kash, and L. H. Poynor, *An evaluation of the personalized model of reacher training. Final report*. Austin, Texas: Research and Development Center for Teacher Education, The University of Texas, 1974.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Student Perception of Teacher Style (#10)

AUTHOR: Bruce W. Tuckman
AUTHOR/PUBLISHER ADDRESS: Copy of the scale may be found in Journal of Educational Research, 1970, (63): 398-400.

CRITERIA EVALUATED: 16

SUBJECT OF OBSERVATION

- ☐ TEACHER AND PUPILS
- ☒ TEACHER ONLY
- ☐ PUPIL ONLY
- ☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

- ☒ 1 ONLY
- ☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

- ☒ LIVE (RETROSPECTIVE RATINGS)
- ☐ LIVE WITH CODING EQUIPMENT
- ☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

- ☐ AFFECTIVE
- ☐ COGNITIVE
- ☒ PROCEDURE OR ROUTINE
- ☐ ACTIVITY
- ☐ SOCIOLOGICAL STRUCTURE
- ☐ OTHER

SETTINGS

- ☒ CLASSROOM, ANY SUBJECT
- ☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

- ☐ CATEGORY CHANGE
- ☐ TOPIC CHANGE
- ☐ SPEAKER CHANGE
- ☐ TIME UNIT
- ☐ TIME SAMPLE
- ☒ RATING SCALE
- ☐ OTHER

OBSERVER(S) REQUIRED

- ☒ 1 PERSON (STUDENTS)
- ☐ TEAM OF 2 PERSONS
- ☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

- ☒ RESEARCH
- ☐ TRAINING
- ☐ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 143-144.

Name: STUDENT PERCEPTION OF TEACHER STYLE

Author: Bruce W. Tuckman

Type of Measure: Rating scale

Availability: A copy of the scale may be found in: Tuckman, B. W., "A technique for the assessment of teacher directiveness." *Journal of Educational Research,* 1970, **63**, 398–400.

Description: The 17-item Student Perception of Teacher Style (SPOTS) rating scale was designed to measure student perception of teacher directiveness. The instrument is based on concepts of directive teaching which were determined by the author from a review of the literature and from classroom observation. These concepts emphasize the structure and quality of interpersonal relationships within junior high and secondary school classrooms. The original SPOTS consisted of 32 items. Two other instruments, the Observer Rating Scale (ORS) and the Teacher Style Checklist (TSC), were developed to refine and determine the construct validity of the SPOTS. Two trained observers rated 22 11th- and 12th-grade teachers on the ORS and the TSC. These teachers were concurrently rated by their students on the SPOTS. Based on intercorrelations between the 3 tests, item analysis, and factor analysis, the original 32 items were reduced to the present 17, yielding a more efficient instrument.

Sample Items: Each SPOTS item consists of a stem followed by a response scale. The student marks one of the 9 points on the scale which best describes his perception of the teacher or classroom. The instrument contains items of the following type:

1. Your instructor is mainly concerned with

1 2 3	4 5 6	7 8 9
The number of facts you know.	Getting an idea across to you.	Your ability to think for yourself.

2. Students in our classroom

1 2 3	4 5 6	7 8 9
Only talk if the instructor asks them a question.	Sometimes volunteer information.	Frequently initiate discussion.

3. For group projects or class committees the teacher

1 2 3	4 5 6	7 8 9
Gives us exact instructions on what to do.	Makes suggestions about the project	Allows the group members to make the decisions.

Reliability: Reliability of the SPOTS was investigated as the instrument was revised (Tuckman, 1970). The mean SPOTS score on each item of the 32-item scale was correlated with the total mean score for each of 22 teachers. Seventy-eight percent of the item scores correlated significantly with total scores, the coefficients ranging from .45 to .91. The 7 items demonstrating the lowest item-total correlations were omitted from the final scale.

Interrater reliability was determined by ranking each student according to the agreement between his SPOTS score and the mean SPOTS score for each of 22 teachers, and then correlating these agreement figures with each teacher's average SPOTS rating. The ratings of students ranking 1 through 4 correlated approximately .95 with the class average; those of students ranking 5 through 8 correlated approximately .84 with the class average; and those of students ranking 9 and 10 correlated .77 and .69, respectively. The ratings of students ranked below 10 did not show high agreement with the class average.

Validity: A significant ($p < .01$) correlation (.53) was obtained between the mean SPOTS rating for each teacher and his corresponding ORS rating. Similarly, a correlation of .31 ($p < .01$) was obtained between the SPOTS mean for each teacher and his mean TSC score. The TSC and ORS correlated .75 ($p < .01$), indicating that they are comparable measures.

Administration and Scoring: The SPOTS is easy to administer. The average high school student should be able to complete the instrument in a short time with minimal instruction. The teacher's rating is obtained by computing the mean score for each item as well as a mean for all of the items combined.

Norms: Item and total means are available on ratings given by 363 male students at 2 vocational/technical high schools to 22 11th- and 12th-grade male teachers, each of whom had at least 5 years of teaching experience.

Comments: Remmers (1963) suggested 5 criteria for evaluating student rating scales, and the Student Perception of Teacher Style appears to meet all 5. The instrument demonstrates sufficient (1) objectivity, (2) reliability, (3) sensitivity in discriminating between instructors, (4) utility and (5) relevance, in this case to the construct of directiveness.

Because the instrument is fairly new, and psychometric data still somewhat limited, its use should be restricted to research at present.

References: Remmers, H. H., "Rating methods in research on teaching." In N. L. Gage (ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.
Tuckman, B. W., "A technique for the assessment of teacher directiveness." *Journal of Educational Research*, 1970, **63**, 395–400.

OBSERVATION INSTRUMENT DESCRIPTION

INSTRUMENT: Teacher Image Questionnaire (#11)

AUTHOR: Roy C. Bryan

AUTHOR/PUBLISHER ADDRESS: Education Feedback Center, Western Michigan University, Kalamson, Michigan, 49001

CRITERIA EVALUATED: One factor analytic study suggests that it is mainly teacher's popularity and charisma which is being rated.

SUBJECT OF OBSERVATION

☐ TEACHER AND PUPILS
☒ TEACHER ONLY
☐ PUPIL ONLY
☐ SMALL GROUPS

NUMBER OF SUBJECTS OBSERVED

☒ 1 ONLY
☐ MORE THAN 2 PEOPLE IN CLASSROOM

COLLECTION METHODS

☒ LIVE (RETROSPECTIVE RATINGS)
☐ LIVE WITH CODING EQUIPMENT
☐ VIDEO AND/OR AUDIO TAPE

CATEGORY DIMENSIONS

☒ AFFECTIVE
☒ COGNITIVE
☒ PROCEDURE OR ROUTINE
☐ ACTIVITY
☐ SOCIOLOGICAL STRUCTURE
☒ OTHER (TEACHER'S CHARISMA)

SETTINGS

☒ CLASSROOM, ANY SUBJECT
☐ CLASSROOM, SPECIFIC SUBJECT

CODING UNITS

☐ CATEGORY CHANGE
☐ TOPIC CHANGE

☐ SPEAKER CHANGE
☐ TIME UNIT
☐ TIME SAMPLE
☒ RATING SCALE
☐ OTHER

OBSERVER(S) REQUIRED

☒ 1 PERSON (STUDENTS)
☐ TEAM OF 2 PERSONS
☐ 2 TEAMS OF 2 PERSONS

USES REPORTED BY AUTHOR

☐ RESEARCH
☐ TRAINING
☒ EVALUATION

REFERENCE:

Evaluating Classroom Instruction: A Sourcebook of Instruments, Gary D. Borich and Susan K. Madden, Addison-Wesley Publishing Company Inc., Don Mills, Ontario, 1977, Pages 145-146.

Name: TEACHER IMAGE QUESTIONNAIRE

Author: Roy C. Bryan

Type of Measure: Rating scale

Availability: Bryan, R. C., *Some observations concerning written student reactions to high school teachers.* Kalamazoo, Michigan: Student Reaction Center, Western Michigan University, 1968. A copy of the instrument may also be obtained from the Education Feedback Center, Western Michigan University, Kalamazoo, Michigan 49001.

Description: The Teacher Image Questionnaire (TIQ) contains 16 questions designed to obtain the opinions of students in grades 7 through 12 about their teachers, as well as 2 optional items which allow students to comment upon the teacher's strengths and weaknesses. Although the TIQ was initially designed to provide feedback for improving teacher effectiveness, it has also been used in studies of teacher image.

This instrument was developed on the theory that there are 3 kinds of criteria for judging teacher effectiveness: product criteria, presage criteria, and process criteria. Since product criteria (long-term outcomes) are difficult to measure, and presage criteria (teacher experience variables) do not provide feedback which can help the teacher improve his or her effectiveness, process criteria were chosen as the basis of the TIQ. Process criteria include variables such as teacher attitude, teacher behavior, and student perceptions of teacher performance. Student perceptions were considered the most important factor in determining teacher effectiveness since they reflect the other 2 process variables, and therefore were selected as the assessment technique for the TIQ.

Sample Item: The TIQ measures the student's reaction to questions about his or her teacher. Responses vary from "poor" to "excellent" on a 5-point scale. Items of the following type are included on the questionnaire:

1. *Knowledge of Subject Matter:* (Does your teacher seem to have a good grasp of the subject matter?)

   poor fair average good excellent

2. *Homework:* (Does your teacher give homework assignments which are interesting and appropriate?)

   poor fair average good excellent

3. *Openness:* (Can your teacher see things from a variety of perspectives?)

   poor fair average good excellent

Reliability: Bryan (1968) reports reliability data collected in 50 randomly selected secondary school classes, each composed of 24 to 32 students. Classes were split

into chance halves, and estimates of internal consistency were computed for the first 12 TIQ items and corrected with the Spearman-Brown formula. These estimates ranged from .82 to .95.

Validity: Bryan (1968b) correlated TIQ ratings completed by students with those completed by administrators on a group of 114 teachers and obtained a coefficient of .61. In a similar study, involving 38 teachers, student and administrator ratings correlated .68.

Bryan (1966) also reported that teacher image, which is normally resistant to change, can be modified with the help of TIQ profiles. In periods up to 2 years, 57% of a group of teachers who had access to their TIQ ratings were able to produce significant gains on 1 or more of the TIQ variables.

A factor analysis of the TIQ (Coats, 1970) produced no meaningful or clearcut item clusters. While student-centered items tended to cluster together, as did structure-related items, one single factor, labeled teacher charisma, accounted for 61.5% of the variance. Coats also found that correlations among items ranged from .32 to .85.

Norms: The TIQ has been administered to students in the 7th through 12th grades. Means and standard deviations computed on the responses of these students are available for 1,630 teachers who were rated on the TIQ from 1970 to 1973.

Administration and Scoring: The TIQ can be administered to an entire class in about 15 minutes. After the instrument has been administered, completed questionnaires are placed in an envelope (along with class and teacher ID forms) and mailed to the Student Reaction Center (Western Michigan University), where answers are analyzed and a report is prepared and sent to the teacher. Confidentiality is assured, and the teacher may request letters of recommendation based on his TIQ profile.

Comments: The TIQ is an inexpensive and useful tool for determining student attitudes toward the teacher. For a nominal fee, the teacher receives a report and profile of his students' reactions. There is some danger, however, that this instrument may measure teacher charisma or popularity rather than actual teaching effectiveness.

References: Bryan, R. C., "The teacher's image is stubbornly stable." *Clearing House*, 1966, **40**, 459–460.

Bryan, R. C., "High school students view classroom control." *Clearing House*, 1968, **42**, 345–346. (a)

Bryan, R. C., *Some observations concerning written student reactions to high school teachers.* Kalamazoo, Michigan: Student Reaction Center, Western Michigan University, 1968. (b)

Coats, W. D., *Students' perceptions of teachers: A factor analytic study.* Paper presented at annual meeting of the American Educational Research Association, Minneapolis, Minnesota, March, 1970.